Statistical analysis of the reliability of complex systems for maintenance planning

Pedersen, Thomas Espelund

Publication date:
2003

Document Version
Publisher's PDF, also known as Version of record

Preface

This thesis is the result of a Ph.D. study entitled "Maintenance and replacement strategies for complex systems". The Ph.D. study was undertaken at Informatics and Mathematical Modelling (IMM) at the Technical University of Denmark (DTU) under the Ph.D. study program for mathematics. The research work has been performed at the Danish Defence Research Establishment (DDRE) which has also financed the Ph.D. study. Chief supervisor has been Professor Poul Thyregod, IMM, with senior researcher Steen Livbjerg and senior researcher Christian Max Møller, DDRE, acting as co-supervisors.

The Ph.D. thesis consists of this overview report together with five other reports. Two of these are included as appendices in this book, while the others, subject to certain restrictions, can be made available by the DDRE.

I would like to thank all the people who have provided advice, help, cooperation and encouragement in the course of this study, particularly Poul Thyregod, Steen Livbjerg, Christian Max Møller, Svend Clausen, Per Garhøj, and Torben Christensen. I would also like to thank everyone at the DDRE, particularly the ORS department, for providing such an encouraging and stimulating environment, and everyone in the KFOR 5 OA Cell for their friendship, cooperation, and inspiration.

Thomas Espelund Pedersen

Danish abstract - Dansk resumé

Denne rapport beskriver en metode til analyse af fejl- og vedligeholdelsesdata fra en population af komplekse reparerbare systemer med henblik på at forbedre effektiviteten af vedligeholdelsen. Rapporten er en del af ph.d. projektet "Vedligeholdelses- og udskiftningsstrategier for komplekse systemer", som har til formål at analysere fejl- og vedligeholdelsesdata med matematiske og statistiske metoder med det formål at forbedre vedligeholdelsesprocedurer i det danske forsvar.

I første del af rapporten introduceres problemet omkring vedligeholdelsesplanlægning og et overblik over modeller for pålidelighed, fejlprocesser og vedligeholdelsesplanlægning. Dette overblik er struktureret så det illustrerer processen at vælge en brugbar model til et givet datasæt, idet der fokuseres på forskellige mål for tid og de forskellige modellers krav til data.

Anden del af rapporten beskriver analysen af to datasæt fra det danske forsvar. Datasættene analyseres med en grafisk metode, Nelson-Aalen plots, så vel som med en multiplikativ intensitetsmodel med proportional intensitets regression, som er en parametrisk model.

Abstract

This report describes a method for analysing failure and maintenance data for a population of complex repairable systems with the aim of improving maintenance efficiency. It is part of a Ph.D. study, titled "Maintenance and replacement strategies for complex systems", the objective of which is to analyze failure and maintenance data using mathematical and statistical models in order to improve maintenance procedures in the Danish Defence.

The first part of the report introduces the maintenance planning problem and presents an overview of models for reliability, failure processes, and maintenance planning. This overview is structured to highlight the process of choosing a proper model for a given data set, focusing on different measures of time and the data requirements for the different models.

The second part of the report describes the analysis of two data sets from the Danish Defence. The data sets are analyzed using a graphical method, Nelson-Aalen plots, as well as multiplicative intensities models with proportional intensities regression, which is a parametric model.

Contents

CHAPTER 1

Introduction

The Danish Defence operates large numbers of increasingly complex and expensive systems such as vehicles, aircraft and ships. These systems are expensive to acquire, operate and maintain, and for this reason, efficient maintenance is of great economic importance. Therefore, the Danish Defence Research Establishment has chosen to focus on maintenance planning as an area of research, and the Ph.D. study described in this report is part of this effort.

1.1 Ph.D. study

The aim of the study, titled "Maintenance and replacement strategies for complex systems", has been to develop a method for analyzing failure and maintenance data using mathematical and statistical models in order to improve maintenance procedures. The Ph.D. study was undertaken at Informatics and Mathematical Modelling (IMM) at the Technical University of Denmark (DTU) under the Ph.D. study program for mathematics. The research work has been performed at the Danish Defence Research Establishment (DDRE) which has also financed the Ph.D. study. Chief supervisor has been Professor Poul Thyregod, IMM, with senior researcher Steen Livbjerg and senior researcher Christian Max Møller, DDRE, acting as co-supervisors. The Ph.D. thesis consists of this overview report together with five other reports. Two of these ([29], [30]) are included as appendices in this book, while the others ([26], [28], and [27]), subject to certain restrictions, can be made available by the DDRE.

1.2 This report

This report has two main parts. In the first, the problem of maintenance planning is described, relevant concepts used throughout the report are defined, and a conceptual model of the problem is developed. Also, the mathematical theory used for modelling failure processes and reliability is described, and an overview of different models from the literature is given. In the second part of this report, the method which has been developed is described, and two examples are given where the method is used on actual data sets from the Danish Defence. Finally, the model is evaluated based on the experience with these two examples.

Chapter 2

Problem definition and conceptual model

Before one applies the tools of mathematical modelling to a practical problem, it is often necessary to set up a conceptual model of the problem, describing the problem verbally and defining the relevant factors, concepts and terms. This section presents such a model for the problem of maintenance planning for a population of complex systems. First some basic terms and concepts are defined, and then the conceptual model, explaining the relations and interactions between the elements of the model, is described. For definitions and terminology, this report borrows heavily from [4].

2.1 Systems and parts

First, it will be helpful to define more clearly what is understood when we talk about systems. At the most basic level, a system can be defined as a collection of parts or components designed to perform one or more specific functions. Hence, a part, unlike a system, can not be disassembled, but a part in a system can be replaced with another similar part or at least a part performing a similar function. A system with a large number of parts, often organized in an architecture of several subsystems, is called a complex system. What constitutes a "large" number of parts is not immediately given, but is in practice decided by the properties of the problem at hand. Simpler systems can be analyzed by looking at the system architecture and the characteristics of the individual parts. In practice, however, some systems have so many parts and a structure so complex that they are impossible to analyze part-by-part, and other methods have to be employed. The systems treated in this

report all belong in the latter category.

2.2 Failure, repair and preventive maintenance

When a system loses the ability to perform one or more of its functions, it is said to experience a failure. This may seem trivial, but in practice it is often hard to specify the exact moment when a system has failed. A system may perform its tasks at a reduced level, for example, and failure will then have to be defined as occurring when the level of performance is no longer tolerable. In the following, failure will be considered a random event, occurring at a specific point in time. This means that failures do not follow any fixed pattern, and it is not possible to know the point in time when a system fails until after the fact.

Alternatively, one may choose to see the condition of the system as a continuous variable. Striking a middle ground between the two, delay time models (see [9]) describe a failure as a two-step process, where a fault occurs, and may be found and corrected, some time before the actual system failure. For the problem and the systems treated here, however, none of these two alternatives are deemed suitable.

There are, of course, many ways in which a system may fail. Methods exist for dealing with this, FMECA (Failure Mode Effects and Criticality Analysis) being the most well known, but these are more geared towards design than analysis of existing systems. In the following, therefore, failure is used as a general term in accordance with the definition above. In the analysis of a given real life problem, information about the actual failure may be taken into account, and general methods can be adapted for this purpose. Theory and methods dealing specifically with different types of failure will not be covered here, however.

Repair is the process of restoring the system to functionality. It is sometimes referred to as corrective maintenance. Systems such as those considered in this report are repairable systems, as opposed to non-repairable systems, which have to be replaced when they fail. An important consequence of a system being repairable is that it may experience more than one failure during its lifetime. Parts are considered non-repairable, and a repairable system can usually be broken down into non-repairable parts. A repair would then involve identifying and replacing broken parts.

Maintenance operations other than repair usually involve some form of disassembly and inspection of the system to identify worn parts or other potential failure risks, followed by action to reduce the risk, such as replacing the parts. This is called preventive maintenance in the following.

2.3 Reliability, availability and maintenance costs

As indicated above, the purpose of preventive maintenance is to reduce the risk of failure, which loosely put is the same as increasing reliability. A more formal definition is that reliability is the probability of no failures in a specific time interval. In practice, a useful definition will often have to be conditioned on more than the

time interval, such as whether the system is in use throughout the interval. Mission reliability is the probability of not suffering any failures while performing a given limited task or mission ([18]).

The availability of a system is defined as the fraction of a given time period where the system is available to perform its designated function; that is, not down for repair or maintenance ([6]). If failures cause downtime, which is usually the case, the availability is a random function. Mission availability is the probability that a system is available when it is needed to perform a function.

In general, maintenance costs are divided into two parts: the price of repairs and the price of preventive maintenance ([6],[3]). Often, the cost per unit of time is used to get comparable results. The cost of repair is incurred whenever a system fails and needs repair, and may include costs suffered by the system failing during operation. Because failures are modelled as random, repair costs will also be random. In contrast, preventive maintenance costs are incurred based on a decision to perform preventive maintenance, and therefore easier to control. Of course, one may still chose to make the decision based on random factors.

As both availability and cost are random, planning for the future usually involves working with the expected availability and expected costs.

2.4 Conceptual model

In managing the maintenance of a fleet of complex, repairable systems, there are two basic objectives to be met: Minimize costs and maximize availability. How these objectives are balanced depends on the individual case, often involving the relationship between procurement and maintenance costs ([20]). Generally, if systems are relatively cheap, it would make sense to buy extra systems to make up for low availability, while, if the opposite is the case, one would spend more on maintenance to attain a high availability. This assumes, of course, that the option of procuring extra systems is available. Alternatively, a fixed lower limit for the availability may be set, or we may search for an efficient frontier, i.e. the set of non-dominated solutions.

The relationship between the level of maintenance and availability is not straightforward, however. As indicated above, the only decision variable in the problem is the procedure for preventive maintenance. The maintenance plan can be formulated in many different ways, but initially it will just be assumed that we can select a higher or lower level of maintenance. It is clear that a high level of maintenance will increase maintenance costs, and it would also be fair at this stage to assume that it will increase reliability. Furthermore, it would be reasonable to assume that it takes more time, and therefore decreases availability, because systems spend more time down for maintenance. Increased reliability, however, means less downtime because of failures, leading to increased availability. These relations are illustrated in figure 2.1.

If procurement of new systems is an option, it is natural to expand the problem to include replacement planning. If it is assumed that maintenance costs rise over time,

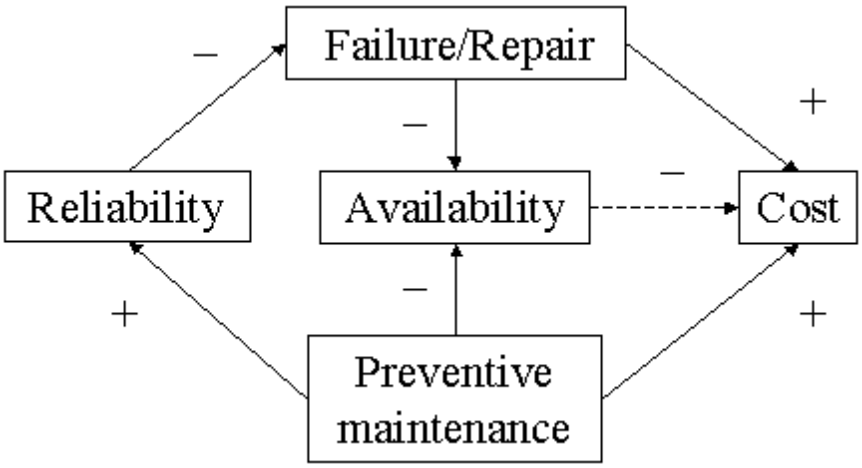

Figure 2.1: Conceptual model.

it will at some point be cheaper to replace a system than to continue maintaining it. While this complicates the problem, especially the cost calculation, it does not change the basic mechanics described above. In fact, replacement can be seen as just another maintenance operation, with a cost equal to the procurement cost and an increased reliability gained by replacing the old system with a new.

2.5 Problem definition and limitation

The simple model of the problem given in section 2.4 clearly illustrates the core problem of reliability centered maintenance planning: preventive maintenance must be performed so that the benefits gained from the preventive maintenance outweigh the price of performing the maintenance ([13], [34]). Note that this applies regardless of whether the objective is to minimize costs, maximize availability or a combination of these.

Simple as this may seem, in practice, a number of problems present themselves. Most have to do with quantifying the effects described in the conceptual model. First and foremost of these is the effect of preventive maintenance on reliability. Reliability itself can only be measured by observing the use and failures of the systems, which is therefore the most important data requirement. The duration of maintenance actions, preventive or repair, will also have to be observed to measure their impact on availability.

Finally, the simplification made initially that maintenance planning is simply a matter of selecting a level of maintenance with a given cost and a given effect clearly does not hold true in reality. For most systems, there is a number of different inspections, tests, and other maintenance operations which can be performed, and

each has its own, often random, cost, duration, and effect on reliability. Also, there is no guarantee of any direct relationship between these characteristics, i.e. it is not given that an expensive maintenance operation will have a larger effect on reliability. Therefore, thoroughly evaluating the effects of preventive maintenance requires large amounts of data from systems operating under different conditions and maintenance plans, ideally spanning the whole space of possible maintenance plans ([32], [35]).

In practice, therefore, the task becomes one of analyzing the available data, engineering knowledge, and information, to gain knowledge about the mechanisms described in the conceptual model, specifically the failure process and the effects of maintenance operations. In addition, knowledge about the cost structures may be needed. The more insight one is able to obtain, the more completely the maintenance may be optimized, and vice versa.

Chapter 3

Mathematical modelling of failure processes and reliability

In section 2.2, failures of a system were introduced as random events occurring at specific points in time. This makes it natural to view the failure process as a stochastic point process. Mathematically, a stochastic point process is described by a series of random variables. There are different ways to express a failure process, and choosing the right way is important, as not all are equally well suited to a given problem.

In this chapter some theory and tools for mathematical modelling of failure processes and reliability will be described. First, an overview of different models and types of models for reliability and failure processes is given. Finally, there is a brief note on mathematical optimization methods for maintenance planning.

3.1 Time and risk

In the following, the term time is used without further comment to measure a system's exposure to risk and the point of failure, for example in connection with times of failure or time between failures. In practice, however, the choice of time scale is an important part of modelling a failure process. The date and time of a failure or the number of hours and minutes on the clock between two failures, called calendar time or real time in the following, may not be the most relevant measure. In many cases, systems are only operating part of the time, and in these cases, it may be more relevant to measure time in a form of operational time, i.e. to "stop the clock" when

the system is not operating, and therefore not at risk of failure. To do this, however, information is needed about when the system is operating.

In fact, though it is used throughout this report and in the literature in general, the word time is too restrictive. For a car, for instance, most people would agree that the number of kilometers on the odometer is at least as important a measure of its condition and the risk of failure as its age or the number of hours it has been operated. This example also illustrates that operational time may be more or less closely linked with calendar time, depending on the pattern of use of the systems.

Other and more complex measures of operational time can be conceived, all depending on the specific situation at hand. It is important to also take the maintenance planning aspect into account, however. If maintenance is planned on a calendar, it makes sense to also perform the analysis in calendar time, accepting a less accurate description of the failure process. Results in some form of operational time may not be useful, unless followed up by an analysis of the relationship between operational and calendar time.

When dealing with multiple systems, it is often helpful to introduce the risk set $R(t)$ at time t, constituted by the systems under risk of failure at a given time t. The risk set is used when determining the overall risk of failure in a group of systems. If the systems are considered similar, it is usually sufficient to consider the size of the risk set, equal to the number of systems at risk at time t, $V(t)$.

3.1.1 Lexis diagrams

Using calendar time instead of operational time may introduce some extra noise in the analysis, but if operating time is evenly distributed over calendar time, it may not make a large difference. As described above, other situations arise where exploring the relationship between calendar and operational time is necessary. To analyze this question further, a so called Lexis diagram[36], plotting operational time against calendar time, can prove quite useful.

In general, the Lexis diagram is used to get an overview of the dynamics of a sample or group of individual subjects. Traditionally, the Lexis diagram has long been used by demographers to plot the age of individuals in a study against calendar time, visualizing sampling patterns in order to predict how these bias the results[23]. For repairable systems, the Lexis diagram may also be used to explore the dynamics between operational time and calendar time. If the Lexis diagrams for the systems approach a straight line (figure 3.1, left), the choice of time scale is not important, whereas large deviations (figure 3.1, right) could imply significant variations if calendar time is used.

With a large number of systems, it can be difficult to get an overall picture of the relationship between the two time scales. In this situation, it can prove useful to produce an "average Lexis diagram", plotting $\left(t_{calendar}, \overline{t_{opr}}\left(t_{calendar}\right)\right)$, where $\overline{t_{opr}}\left(t_{calendar}\right)$ is the average operational age of all systems active at calendar time $t_{calendar}$. Additionally, it is useful to plot $\left(t_{calendar}, \overline{t_{opr}}\left(t_{calendar}\right) \pm 2\widehat{\sigma_{t_{opr}}}\left(t_{calendar}\right)\right)$, where $\widehat{\sigma_{t_{opr}}}\left(t_{calendar}\right)$ is the estimated standard deviation of the operational age of

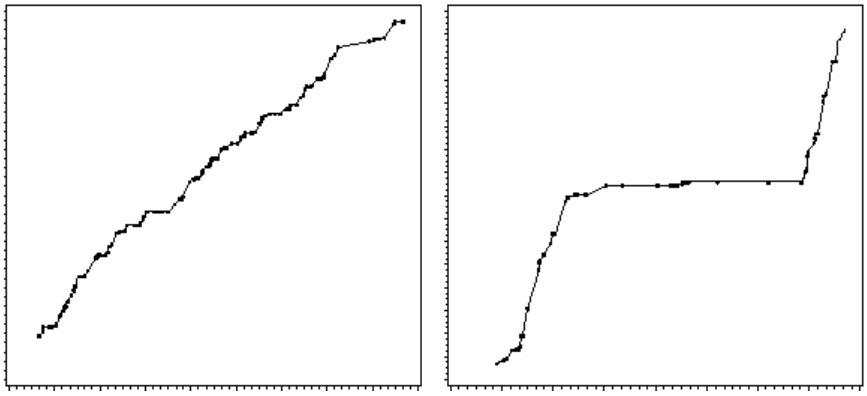

Figure 3.1: Examples of Lexis diagrams.

all systems active at calendar time $t_{calendar}$. Together, these plots can serve to illustrate the connection between the time scales, with a narrow band around the average suggesting a close correlation. An example is shown on figure 4.2, page 28.

3.2 Local time

One way to describe a failure process is through the times between failure, also called interarrival times, where the amount of time from one failure to the next is a random variable $X \in \Re^+$. Often, some extra information is included by ordering the failures so that $X_i \in \Re^+$ indicates the time between the $i-1^{st}$ and the i^{th} failure. This information is very important in order to make the correct choices when selecting the model to use (See section 3.4). [5] describes some errors which may arise if the chronological order of the observations is ignored.

The time scale measuring time between failures is sometimes called local time. ([4] attributes this term to Jewell.) An important underlying assumption made in using local time is that something significant happens to the system, and the reliability of the system, as it fails and is subsequently repaired. Usually, it is assumed to be restored to its initial reliability (Renewal process, see section 3.2.2), and this type of model is therefore often referred to as good-as-new models or perfect repair models.

In addition to the lengths of the intervals between failures X_i, more information about the system, environmental factors etc. may be contained in a covariate vector Z_i.

3.2.1 Reliability measure

The most obvious way to express system reliability in local time is through the probability distribution of X, given either as the distribution function $F(x)$ or the density function $f(x)$. Sometimes, $R(x) = 1 - F(x)$ is called the reliability function.

However, the force of mortality (FOM) or hazard rate $h(x) \equiv \frac{f(x)}{1-F(x)}$ often proves more useful. $h(x)$, or more accurately $h(x)\,dx$, can be interpreted as the probability of failure in the interval $[x; x+dx]$, given that the failure has not happened yet at time x. This condition is important, because in local time, only the next occurring failure is important.

3.2.2 Models

The simplest class of models in local time is the renewal process, where all interarrival times X_i are independent and identically distributed (IID). A special case of the renewal process is the homogeneous Poisson process (HPP), where all X_i follow the same exponential distribution. This makes the HPP a renewal process with constant FOM. See section 3.3.2 for further treatment of the HPP.

More complex variations of the renewal process include the superimposed renewal process (SRP) and the branching renewal process (BRP). The SRP is generated by a union of two or more independent renewal processes, so that a failure in any of the renewal processes is counted as a failure in the SRP. In the BRP, a primary renewal process at each failure generates a secondary renewal process, which continues to run superimposed with the primary and any previously generated processes. To keep the number of failures from increasing boundlessly, the secondary processes are usually assumed to be finite, i.e. they stop after a number of failures.

In [11], Cox introduced the proportional hazards model, sometimes called Cox regression, where the FOM depends on covariates as well as time, taking the form $h_i(x_i, \mathbf{z}_i) = \lambda_0(x)\exp(\mathbf{z}\boldsymbol{\beta})$. $\lambda_0(x)$ is a baseline FOM, and $\boldsymbol{\beta}$ is a parameter vector for the covariates. Using the previous failure history of the system as covariates, a full history model is developed in [19]. Parameter estimation in the proportional hazards model using maximum partial likelihood is further developed in [12].

Models with different distributions $F_i(x_i)$ and FOM $h_i(x_i)$ depending on the number i of the next failure may be formulated as Markov models, making the convenient assumption that X_i depends only on the present state $i-1$. Other Markov models jump back and forth between a finite number of states, usually just the two states working and failed. Markov models of failure processes are treated further in [24].

3.3 Global time

Of course, the alternative to local time is to describe the process in global time. Here, the random variable of interest is $T_i \in \Re^+$; the arrival time for the i^{th} failure, which is measured from an origin $T_0 \equiv 0$, rather than from the previous failure. The important difference between the two measures is underlined by using X_i for local time and T_i for global time. The relation between recordings in the two time-regimes is given by $T_i = \sum_{i=1}^{j=i} X_j$ and $X_i = T_{i-1} - T_i$.

In general, calculating the distribution function for T_i from that of the X_is is a complex task, if at all possible, because it involves i fold convolution integrals of the

distribution functions of the X_is. A much more useful way to model the process is to describe the counting process $N(t)$ which counts the number of failures experienced by the system in the (global time) interval $[0;t]$.

Unlike local time, the global time scale is not affected by the occurrence of failures. Modelling the reliability of a system in global time therefore carries with it the assumption that the reliability is not affected by failure and repair. In other words, the repaired system is good-as-old (or bad-as-old), because the reliability is the same after a repair as it was just before the failure. Therefore, it is also known as minimal repair.

As in local time, a covariate vector $Z(t)$ may hold additional information. Note that in global time, we can allow the covariates to change over time. Also, it is useful to define a risk function $Y(t) \in \{0,1\}$, indicating whether a system is at risk at time t ($Y(t) = 1$) or not ($Y(t) = 0$). All observations about the process up to time t; failures, covariate values, and the risk function, is collectively called the filtration or history of the process, denoted $\mathcal{F}(t)$.

3.3.1 Reliability measure

Because more than one failure can be considered in global time, the condition made in the definition of the FOM becomes irrelevant. As already mentioned, probability distributions for T_i are also hard to come by. By way of the counting process, however, a useful measure can be defined.

For a counting process $N(t)$ there exists a compensator $\Lambda(t)$ so that $N(t) = \Lambda(t) + M(t)$ where $M(t)$ is a martingale, meaning, in essence, a pure noise process. The martingale property may be written as

$$E\left[dM(t) \mid \mathcal{F}\left(t^{-}\right)\right] = 0 \tag{3.1}$$

where $\mathcal{F}(t^{-})$ is the history, that is, all observations up to time t. The intensity process $\lambda(t)$ for the counting process $N(t)$ is defined by

$$E\left[dN(t) \mid \mathcal{F}\left(t^{-}\right)\right] = \lambda(t)\,dt \tag{3.2}$$

Because

$$\Lambda(t) = \int_0^t \lambda(u)\,du \tag{3.3}$$

the compensator $\Lambda(t)$ is sometimes called the integrated or cumulated intensity. The theory of counting processes, martingales and intensities is extensively covered in [2].

In [4], the peril rate or ROCOF (rate of occurrence of failures) $\rho(t)$ is defined as

$$\rho(t) \equiv V'(t) \text{ where } V(t) \equiv E[N(t)] \tag{3.4}$$

The main difference between the two is the conditioning on the history of the process. This means that, because the history of the process contains random elements, the intensity process at time t is random, at least until the process has been observed up to t^{-}, whereas the ROCOF is fixed.

3.3.2 Models

The most common class of models in global time is the Poisson process. These models for a counting process $N(t)$ assume that $N(0) = 0$ and that $N(t)$ has independent increments. A process has independent increments if, for all $0 < t_1 < \ldots < t_k$, $k = 2, 3, \ldots$, $N(t_1) - N(0), \ldots, N(t_k) - N(t_{k-1})$ are independent random variables, or in words, that the number of failures in any interval is independent of the previous failure history. The most general form of Poisson process is the nonhomogeneous Poisson process (NHPP). Here, the number of failures in an interval $[t_1, t_2]$ has a Poisson distribution with mean $\int_{t_1}^{t_2} \rho(t)\, dt$, where $\rho(t)$ is the ROCOF. Two widely used parametric forms for $\rho(t)$ are the log-linear ROCOF $\rho(t) = \exp(\alpha_0 + \alpha_1 t)$ and the power law process $\rho(t) = \lambda\beta t^{\beta-1}$, both of which have the advantage that maximum likelihood estimates for the parameters can be found ([4], [10]).

An important special case is the homogeneous Poisson process (HPP), where $\rho(t)$ is a constant. An alternative definition of the HPP is given in section 3.2.2, and in fact the HPP is the one Poisson process that can be applied in both local and global time. In all other cases, the interarrival times in an NHPP are not IID and therefore it is impossible to define an equivalent renewal process.

The risk function $Y(t)$ can be taken into account by using a multiplicative intensity model $\lambda(t) = Y(t)\,\alpha(t)$, where $\alpha(t)$ is the failure intensity for the system, given that it is under risk at time t. Equivalent to the proportional hazards model in local time (see section 3.2.2), a proportional intensities model $\lambda(t) = \lambda_0(t) \exp(\mathbf{z}(t)\,\boldsymbol{\beta})$ may be used in global time. $\lambda_0(x)$ is a baseline intensity, and $\boldsymbol{\beta}$ is a parameter vector for the covariates. A model along these lines was introduced in [33], while in [2] a general model of the form $\lambda(t) = Y(t)\,\alpha(t) \exp(\mathbf{z}(t)\,\boldsymbol{\beta})$, combining multiplicative intensity and proportional intensity regression, is developed. Tests for linear hypotheses in $\boldsymbol{\beta}$ are described in [30].

3.4 Preliminary analysis and model selection

Two main issues should be considered when selecting which model to use when analyzing a set of failure data for maintenance planning: First, the basic assumptions behind the model must hold, and second, the model should be able to provide the type of information needed for solving the planning problem.

The choice should therefore be based on a preliminary analysis of the data, as well as general knowledge and understanding of the systems and the problem at hand. In the literature, there is a tendency to emphasize the first while ignoring the second. [4], [8], and [20], among others, give algorithms for analyzing the data and selecting an appropriate model in such a way as to suggest that this process might be more or less automated (See figure 3.2). In practice however, this is rarely the case, and characteristically, neither mentions problems such as those addressed in section 3.1.

The model selection algorithms are, however, very useful in getting an overview of the conditions that the data need to fulfill for the various models to be applicable. As shown in figure 3.2, it is important to investigate whether there is a trend in

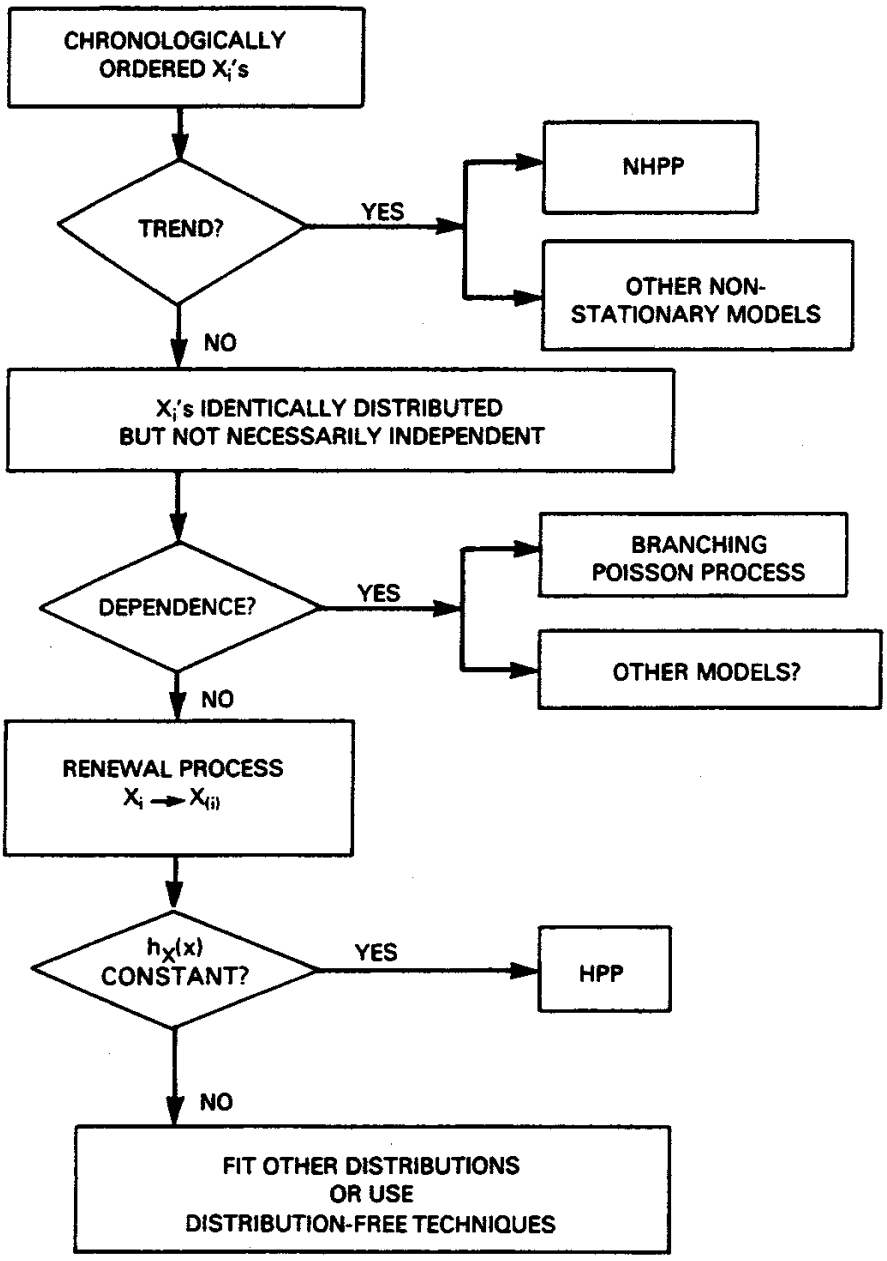

Figure 3.2: Algorithm for model selection. From [4]

the data, i.e. if times between failure become longer or shorter with time. Model selection aside, this question is also important in the overall maintenance planning context. Methods to do this include statistical tests such as the Laplace test (see [4] for an overview of such tests), graphical methods such as the Nelson-Aalen plot (see section 3.4.1 below), and TTT (Total Time on Test) methods (see [21]).

If no trend is found, local time models may be applied, and tests for dependence and distribution may then be applied to determine which model to use. At this point, however, it is worth remembering that the absence of a clear trend does not preclude the use of global time models like NHPP models. There may be practical reasons for doing this, for instance if one wishes to include covariates which change over time in the model, or if, for planning purposes, results in global time are needed.

Also, knowledge about the system may lead one to prefer one time scale or the other. Most importantly, it may be used to consider the good-as-new or bad-as-old assumption which is implied by the choice of local time or global time. If maintenance operations involve major changes to the system, the good-as-new assumption may be appropriate. The replacement of a small number of parts in a complex system consisting of a vast number of parts, however, hardly seems to justify an assumption that the system has been renewed. Of course, the bad-as-old assumption that the system is exactly as before is not entirely correct either, but it is the better approximation to the actual conditions.

Various models to have been suggested to strike the middle ground between the two: imperfect repair, see for example [15] and [22]. Usually, these take the form of a definition of an operational time scale in global time (called virtual age in [22]), conditioned on failure times, so that time is decreased upon maintenance, and the system is "rejuvenated". This is problematic, however, because the resulting time scale is difficult to work with for planning purposes. It is difficult to translate into real time, and plans into the future have to be conditioned on the failure history.

3.4.1 Nelson-Aalen plots

The Nelson-Aalen plot is a non-parametric, graphical method for analyzing the failure process for a group of repairable systems. It was introduced as a tool for analyzing repairable systems in [25], though its statistical properties are already discussed in [1]. For local time, it is analogous to the Kaplan-Meier estimator, described in [17].

The Nelson-Aalen estimator estimates the mean cumulated number of failures per system, MCNF$(t) = M(t)$, defined as $M(t) = E[N(t)]$. The significance of $M(t)$ is due to the fact that the slope of the curve, $m(t) = \frac{dM(t)}{dt}$, gives the intensity or ROCOF at time t. The estimator is defined as follows

$$\widehat{M}(t) = \sum_{s \leq t} \frac{\Delta N(s)}{V(s)} \tag{3.5}$$

where $V(t)$ is the number at risk, i.e. the number of systems active, at age t, and

$\Delta N(t)$ is defined as

$$\Delta N(t) = \begin{cases} 1, & \text{if } t \in \{T_i\} \\ 0, & \text{otherwise} \end{cases}$$

From the definition, it is clear that the Nelson-Aalen plot is an increasing step-function, which jumps only at the times of failure. Therefore, it is sufficient to know the set of jump points

$$\{T_i, \widehat{M}(T_i)\}$$

in order to produce the plot. These points can be calculated recursively using the formula

$$\widehat{M}(T_0) = 0$$
$$\widehat{M}(T_i) = \widehat{M}(T_{i-1}) + \frac{1}{V(T_i)}, i \geq 1 \tag{3.6}$$

Some properties of the Nelson-Aalen estimator are investigated in [31]/[29]. Specifically, these have to do with the bias caused by variations in $V(t)$, especially when $V(t)$ is small. Typically, $V(t)$ is small and increasing at the beginning of the observation interval and decreasing towards 0 at the end. Often, $\widehat{M}(t)$ shows a bias in connection with these developments in $V(t)$, particularly at the end of the interval, where a strong upward bias can be observed. This could lead to erroneous conclusions about a sudden increase in failure intensity, but is shown to be caused by the properties of the estimator.

A thorough treatment of the Nelson-Aalen estimator and its statistical properties can be found in [2].

3.5 Optimization methods

Various methods for finding optimal maintenance plans, replacement times, dimensioning of maintenance facilities, etc. have been put forward in the literature. These methods can be divided into two different classes.

A large number take the form of a closed formula for maintenance costs or the optimal decision, and these are mostly solutions to very specific problems. In developing these, strong assumptions about the failure process, the effect of maintenance, and cost structures, are often necessary. Examples are [7], which presents some quite general solutions for replacement problems with age-dependent rewards and failure rates, [10], where optimal replacement policies when the failure process follows the power law or log-linear NHPP are given, and [32]. Other typical examples are queue models used for dimensioning maintenance facilities, where customers are systems needing maintenance, and servers are maintenance facilities.

Some methods take a more general approach, and may therefore be adapted for wider range of failure processes and cost structures. Stochastic simulation is one such

method, used in [16] and [19] to predict future behavior and optimize maintenance accordingly. Heuristic and metaheuristic methods may also be employed, as for instance in [14], where a genetic algorithm is employed.

For the methods tailored to very specific situations, the problem of course is that in an actual situation, all the assumptions made in developing the method may not hold. This is aggravated by the fact that the assumptions in some cases seem to be made to facilitate the calculations more than to approximate real life situations.

More general methods overcome this problem, but still suffer from the problems discussed in section 2.5: they require detailed and accurate knowledge about the failure process and how it is affected by different maintenance operations and other factors, cost structures etc. This in turn requires massive amounts of data to estimate the relevant parameters. Also, even with the computing power available today, many maintenance planning problems involve systems, cost structures, etc. so complex that the optimization model, even if it is possible to set up, becomes impossible to solve in practice.

The problems of parameter estimation and model verification are usually ignored when optimization methods are presented, but have to be considered for the methods to be used in real life.

Chapter 4

Analysis of Danish defence data

This chapter describes two case studies where maintenance data from the Danish defence are analyzed in order to improve maintenance planning. The first data set is from the army's fleet of Leopard 1A5 tanks, and the second set is from the air force's F-16 fighter aircraft. The two data sets are analyzed using the same basic method, which was developed in the course of working with the first data set. The analysis of the second data set, which has many similarities with the first data set, therefore has the added objective of testing the method.

First, the method used in the analyses will be described and discussed. Then follows a summary of each of the two analyses, and finally a conclusion and evaluation of the method.

4.1 Method selection

The two overall situations covered here, as well as the systems in question, have a number of features in common. This ultimately means that the same method may be used on both, because the selection of tools and models follows a common track.

4.1.1 Maintenance planning problem

The situation in both cases is that economic, political, and other considerations preclude the procurement of new systems. Therefore, the number of systems is given, and replacements need not be considered. In both cases, the systems are maintained by a large organization which also maintains other types of systems, to a large degree

after common guidelines. This greatly restricts the degree to which changes to the overall maintenance practices and organization can be considered, at least based on these analyses alone.

Thus, the focus of the analysis will be on investigating the failure process, including the possible effects of preventive maintenance and other factors. This will be done with a view to providing advice to improve maintenance efficiency, increasing availability and/or reducing costs, within the framework of the existing maintenance organization. Optimization of maintenance operations will not be attempted, for the reasons already stated in sections 2.5 and 3.5, and because of a lack of data with sufficient quality and variation of system histories.

4.1.2 Systems

The systems themselves are extremely complex, each consisting of many subsystems and thousands of parts. This practically rules out any modelling of individual parts or of the system architecture. First, the result would be so complex as to make it intractable and second, it would require the estimation of many parameters with each estimate based on too few observations.

Also, as this is an initial analysis of the problem, the information obtained from modelling the systems on a more detailed level is not necessarily relevant. While information about the deeper causes of failures and particularly vulnerable parts or subsystems is certainly valuable, the object of this analysis is to get a broader picture of the problem. Therefore, detailed modelling of the systems should be part of a later analysis based on the results of this analysis.

Instead, the systems will have to be modelled as complex systems, where the failure process of the system as a whole is considered, without regard to how a failure occurs and which parts are involved in failure and repair.

These considerations, together with the fact that the systems are repairable, point to global time models as the most appropriate. Maintenance planning is also based on global time, meaning that preventive maintenance is performed at fixed intervals in global calendar or operational time, not at times since last failure. This makes it even more attractive to have results in global time.

Besides the information on failures and maintenance operations, the data contain some additional information about the systems and their working environment. As we want to investigate whether these factors influence the failure process, a model which incorporates covariate information will be needed. This also lets us incorporate some, though by no means all, of the history of the systems and their parts as covariates.

4.1.3 Method selection

The exploratory nature of the required analysis means that a graphic method such as the Nelson-Aalen plot (section 3.4.1) is well suited for the initial analysis of the data. The benefits of this approach is that, as the Nelson-Aalen plot is a non-parametric method, very few assumptions have to be made in order to use it. The

main purpose of producing the plot is to get an estimate of the failure intensity and its development over time, for instance to see whether reliability is increasing or decreasing. In addition, producing plots for different measures of time, i.e. calendar and operational time, can help determine the best choice of time scale. Finally, by producing plots for different values of covariates, an initial estimate of the effect of these may be obtained, and assumptions necessary for further modelling, e.g. proportional intensities, may be validated. The choice of time scale may also be aided by producing Lexis diagrams (3.1.1) to illustrate the relationship between time scales.

Covariates which take continuous values can not immediately be illustrated in a Nelson-Aalen plot, particularly if they change over time. Furthermore, we want to quantify the effects of covariates, and we want to test whether these effects are statistically significant. This may be accomplished by fitting a proportional intensities model.

Together, the results from the Nelson-Aalen plots and the proportional intensities model provide at least an initial picture of some key factors in the maintenance planning problem. This in itself may provide some clues on how to improve maintenance efficiency. Also, data requirements for further analysis can be identified.

As we are dealing with global time data from repairable systems, it may well be possible to fit a parametric NHPP model to the data. At this stage, however, this would be of little practical use. For optimization or other purposes, where there is a need to predict future behavior, a parametric model would be beneficial, but for the purposes stated above, the Nelson-Aalen plot will provide sufficient information about the failure intensity.

4.2 Analysis of tank data

The method described above has been applied to a set of maintenance data from 230 Danish Leopard 1A5 tanks. Data has been recorded over a period of several years, covering the entire period where the tanks have been in use in the Danish Army. The graphic analysis, together with a description of the data and a discussion of the choice of time scale, is presented in [26], and the parametric analysis using proportional intensities is presented in [28]. Some interesting properties of the Nelson-Aalen plot, which were observed during the analysis, are presented in [29].

4.2.1 The data

The individual observations in the data are work orders. A work order is produced when maintenance work is performed on a tank. The data come from workshops at a high level in the maintenance organization, meaning that minor repairs and inspections are not included, because these are handled at a lower level.

Each work order contains the following information:

- Identity of the tank.

- Dates for the opening of the work order, the beginning and end of the work, and the return of the tank to the unit.
- Odometer reading at the opening of the work order.
- Description of work performed, identifying the subsystem (weapons systems, engine, or electronics), and the work performed; given by a code for a specific area within the subsystem and numeric codes for specific tasks. Each work order can contain up to four numeric codes for repairs (all in the same area) and up to two codes for inspections.
- Production year of the tank.
- Unit to which the tank is attached.

Work orders concerning only inspections and no repairs are filtered out, since these do not constitute a failure.

Because some work orders for the same tank overlap, the concept of super repairs is introduced. A super repair is the set of work orders for the same tank which overlap with at least one other member of the super repair and none outside the super repair. Thus, the duration of a super repair is from the beginning of the first work order to the end of the last work order.

The overlaps may be the result of repairs involving more than one area in the system, necessitating more than one work order for the same repair, or a repair may have been divided between different workshops. Therefore, it may be reasonable to see the number of super repairs as the true measure of the number of failures. Alternatively, some super repairs appear to be the result of faulty registration of dates, and working with super repairs instead of work orders may therefore introduce extra noise in the data.

Mistakes made in entering the data is one of several problems with the data. Data are recorded for accounting purposes rather than reliability analysis, and this means that work orders are sometimes opened and closed for administrative reasons, and that the dates given do not always reflect the true time when a tank has failed or when it is returned to active duty. Information about when a tank is first put into active service, and when it is transferred to another unit, is also not available, and has to be inferred from the maintenance data, leading to inaccuracies.

4.2.2 Time scale and risk

Several different time scales may be used in the analysis. The simplest is calendar time, where a point in time equals the date, and intervals are measured as the number of days between start and finish. Because the whole history of each tank is observed, the age of the tank, i.e. the time passed since it began operating, can be used as an operational time scale. With operational time, there is the added option of "stopping the clock" while a system is undergoing maintenance.

An obvious alternative operational time would be the odometer reading, but as these are only registered sporadically and mostly left blank, this is not a realistic option.

Assumptions also have to be made about whether a system is at risk of failure during repair, or whether it should be removed from the risk set during repair. Of course, not all combinations of time scale, risk during repair, and data points are possible. Per definition, a new super repair can not happen for a tank while one is already in effect. Therefore, it does not make sense to assume that failures can happen during repair when working with super repairs. Conversely, it is clear from the data that new work orders do appear while one is already in effect, so here, assuming no risk of failure during repair is plainly wrong. Furthermore, it does not make sense to stop time during repairs. These considerations leave five logically possible combinations, shown in the table below.

Time scale:	Risk of failure:	Data:
Calendar time	Not during repair	Super repairs
Calendar time	Also during repair	Work orders
Operational, pass during repair	Also during repair	Work orders
Operational, stop during repair	Not during repair	Super repairs
Operational, pass during repair	Not during repair	Super repairs

Nelson-Aalen plots have been produced for each of these combinations. These show that the main difference in results comes with the choice of calendar time or operational time, while super repairs and work orders show very similar results. Therefore, and to avoid the added noise and complexity that comes from working with super repairs, the further analyses are carried out in both calendar and operational time, but on work orders only.

4.2.3 Analysis and results

To illustrate the difference in failure intensity between different units and production years, Nelson-Aalen plots have been produced for different values of these covariates. These show large differences in failure intensity, both for different production years and for different units. Results are fairly consistent whether one looks at calendar time or operational time.

Though the plots appear to indicate some changes in failure intensity over time, these are quite contradictory. Specifically, we see a wear out effect in operational time, with failure intensity increasing as the tanks age, while the opposite seems to be the case in calendar time. These effects can be explained as being the effects of changes in the risk set, however, leading to the conclusion that the failure intensity is in fact fairly constant over time.

There is a large multicollinearity between the two covariates, meaning that tanks from certain years are concentrated in certain units, not evenly spread out among them. Even though this does not explain all the difference between units, it is hard to know from the Nelson-Aalen plots how much variation is caused by production year and how much by unit.

Analyzing the data further using proportional intensities regression sheds some more light on this. The analysis was carried out on three data sets of work orders. One uses calendar time, while the other two use operational time, the difference being that repairs carried out in connection with inspections are filtered out from one of them.

For each data set, models have been fitted with unit, production year, and both unit and year as covariates. A new covariate, the time since last inspection, has also been included, both alongside unit and unit and year combined. The model has been implemented using the SAS procedure PHREG. The covariates have been coded as sets of (0,1) variables, so that in each observation, one variable is 1 and all the others are 0. One unit and one year are chosen as the baseline (all variables are 0), against which the others are measured relatively.

Both unit and year are found to be statistically significant when used alone as covariates in the model. When both unit and year are included, the effects of multicollinearity become visible: parameter estimates change, particularly for the years, where some years do not come out as significantly different from the baseline. The unit parameters, on the other hand, were not very different from what was obtained with the unit as the only covariate. This leads to the conclusion that, though the year may have some effect, the main factor influencing the failure intensity of the tanks is the unit to which they are assigned.

No major differences were found between the results in calendar time and operational time. Removing repairs performed in connection with inspections does make a difference in the parameters for the units, however. This indicates some differences in maintenance practice between the units. Figure 4.1 shows the relative failure intensities for the units in the three data sets. It is worth noting that the three units with the highest failure intensity are armored regiments, while the others are mostly armored infantry. This points to the way the tanks are operated as a factor influencing reliability. Time since last inspection, while statistically significant, does not seem to have any practical influence on the failure intensity.

Based on the results of these analyses, a number of conclusions and recommendations can be made about the maintenance of the tanks. A constant failure intensity and the relatively small effect of the time since last inspection indicate that the tanks are not being worn out. This means that the level of preventive maintenance is at least adequate, and possibly excessive. Lowering the level could have advantages in terms of savings and increased availability, but these can not be quantified based on the available data.

Significant differences have been found between units of different type but also between units of the same type. Further investigations, focused on finding the causes of these differences, could generate useful knowledge about the tanks, and possibly

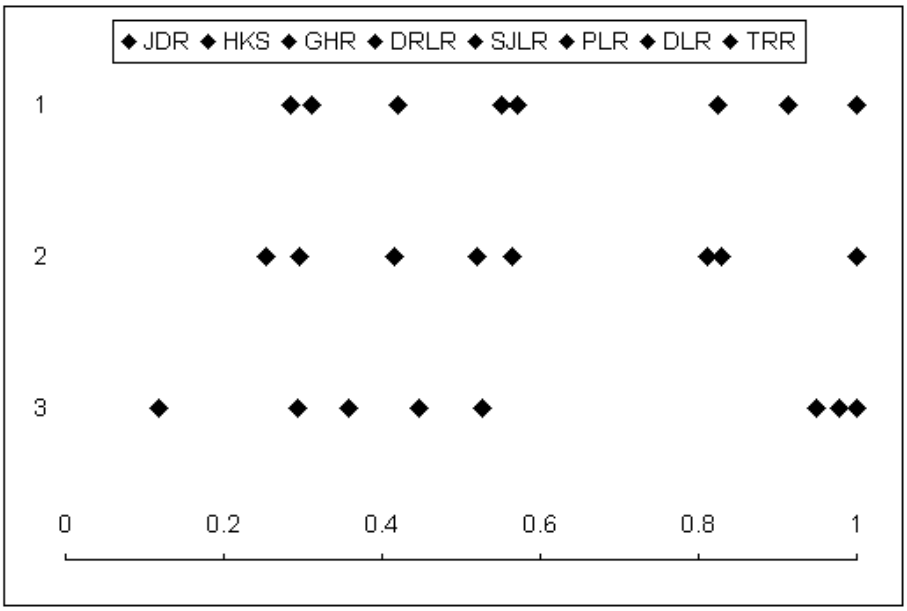

Figure 4.1: Relative failure intensities for the units in the three data sets.

reveal ways to improve maintenance efficiency. At any rate, it would make sense to take the differences into account when maintenance budgets are drawn up.

4.3 Analysis of aircraft data

The method described in section 4.1.3 was developed for the analysis of tank data as described in section 4.2. To test the applicability of the method to a quite similar problem, it has been applied to maintenance data from 70 F-16 fighter aircraft. Again, data has been recorded over several years, though not the entire history of F-16 in the air force. The analysis is presented in detail in [27].

4.3.1 The data

The available data is actually stored in two separate data sets. One set contains records of failures and repairs, while the other contains records of each flight of the aircraft. The two sets are merged into a single data set for the analysis. From the records of the failure data set, the following information was used:

- Identity of the aircraft.
- Code for the circumstances under which the failure was discovered. (Before flight, during flight, mission abandoned, mission continued, etc.)

- Date and time of report.

From the flight reports, the following information was used:

- Identity of the aircraft.
- Squadron to which the aircraft was attached at the time of flight.
- Duration of the flight.
- The aircraft's total flight time to date.

In addition, information has been obtained about which production block each aircraft belongs to.

Total flight time to date is also recorded with the failure data, but because of a high rate of errors in these recordings, the total flight time to date for the aircraft at the time of failure was calculated from the flight reports. The flight reports were also used to determine the squadron attachment history of the aircraft.

The fact that the date and time of the report are used in place of the failure time inevitably causes some inaccuracies. These are hard to quantify, but are generally not significant, because repairs are usually started almost immediately. A single failure can result in several repair actions, but unlike the tank data, these are recorded under a common case number. This, in principle, ensures that failures are only counted once. In reality though, there seems to be some difference in reporting practice, and errors can occur so that failures are counted more than once. Again, the effect of this is hard to quantify, but it should not be a major problem.

These problems aside, data are very carefully collected and accurate. No or very few observations are likely to be missing, which is crucial in a survey such as this, where the entire history of the systems is used in the analysis.

4.3.2 Time scales, risk, and failure

From the information included in the data, it is clear that we have the option of using flight time as an operational time scale. This is highly relevant in terms of planning and operational evaluation, and it is a widely used measure. Other operational time scales might be envisioned, but their practical use is limited by a lack of data. Calendar time is therefore the only other time scale which may be relevant. Hence, calendar time and flight time are the time scales in which the analysis will be performed.

Lexis diagrams for the aircraft and plots of the size of the risk set as a function of calendar time and operational time combine to show that the period of calendar time for which the data have been recorded also covers a fairly narrow band of operational time. Figure 4.2 shows an average Lexis diagram (see section 3.1.1) which illustrates this. In other words, all the aircraft have been and continue to be flying roughly the same number of hours per month or per year, so there is a large correlation between calendar time and operational time. Aircraft purchased later than the bulk of the

fleet to replace losses etc. and hence entering the data later with flight time 0, are an exception from this, and the effect is clearly visible in figure 4.2.

It is natural to assume that the aircraft are under constant risk failure in operational time. In calendar time, however, there will be periods in the history of an aircraft where one may assume that it is not under risk of failure. This could, for instance, be the case during major maintenance operations. Again, because of lack of data to identify such periods, it has to be assumed nonetheless that the aircraft are always at risk of failure. This approximation is relatively safe, however, because the periods in which an aircraft is not under risk are short compared to the total time in which the aircraft has been observed.

In collecting the data, failure is defined quite broadly, including parts changed as part of normal maintenance. Because we are more interested in failures which have a real operational impact, only failures which have been found during missions and have led to the mission being cancelled are considered.

4.3.3 Analysis and results

The covariates which have been analyzed for their possible effect on the failure intensity are: Squadron, production block, and model (one or two seat). All the analyses have been carried out in both calendar time and flight time. The Nelson-Aalen plots produced as the first part of the analysis, for all the aircraft as well as divided by each of the covariates, show some interesting properties.

The failure intensity seems to be, in practice, constant over time; for the aircraft as a whole and for the subgroups, and in both time scales. There are variations, but these can be explained by changes in the risk set and inaccuracies in the data. The most significant differences in failure intensity are found between squadrons, but there are also clear differences between different blocks and models. Muliticollinearity is less of a problem here than with the tanks, as the different models and blocks are more evenly spread out among the squadrons. The differences between the plots suggest that the effects of the covariates can be modelled with a proportional intensities model.

Various models, with some or all of the covariates included, have been fitted to the data in both time scales. As was done with the tank data, the covariate information has been coded as (0,1) variables.

Taken one at a time, the effects of all the covariates are statistically highly significant. The largest differences are found between squadrons, and though these are less dramatic than what was found for the tank units, the failure intensity for some squadrons is more than twice that of others. The intensity for two seat aircraft is also appreciably higher than for single seat aircraft, while the difference between blocks, though still statistically significant, is too small numerically to have much operational impact.

When all covariates are included in the model, multicollinearity between the covariates becomes apparent: The block effect practically disappears, and the model effect is much smaller. This confirms that the squadron is by far the most important

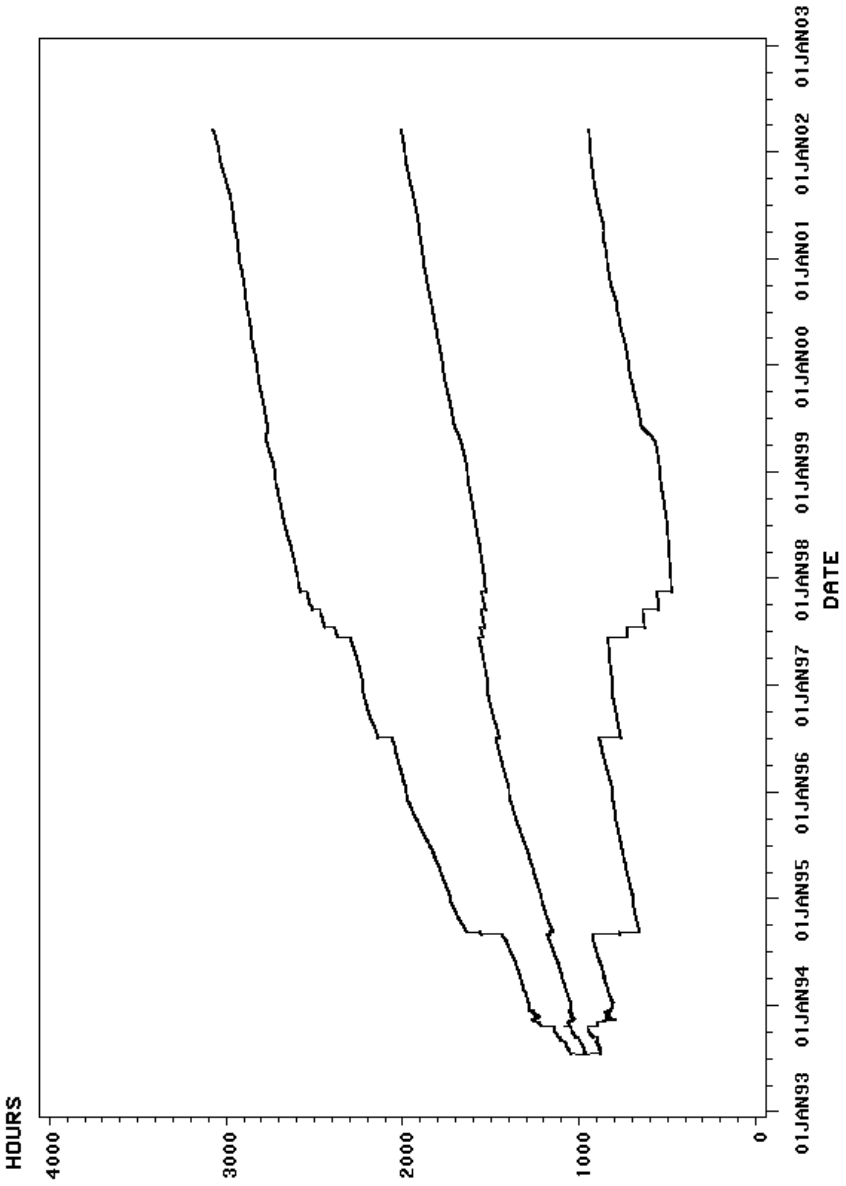

Figure 4.2: Average Lexis diagram $\pm 2\widehat{\sigma}$.

effect on the failure intensity.

Differences in failure intensity are generally higher in operational time, again indicating that the operational environment has more impact on reliability than differences in construction. Even the model effect may be explained by differences in the way two seat aircraft are used. As was the case with the tanks, further investigations into the precise causes of the differences in failure intensity could yield ways to improve reliability and efficiency.

CHAPTER 5

Conclusion

5.1 Reliability modelling and maintenance planning

The literature covering the statistical and mathematical modelling of reliability and failure processes is very extensive. However, it deals almost exclusively with situations where data are given and the questions about local or global time and the choice of time scale have already been dealt with. Indeed, some confusion still exists about the fundamental differences between local and global time, and the importance of choosing a proper time scale is almost completely ignored.

In this report, an effort has been made to present existing models in a context which clarifies the process of choosing the right model for the problem and the data at hand. This is done by focusing on the function of mathematical modelling in the overall problem solving process. To this end, a conceptual model of the maintenance planning problem (Figure 2.1, page 6) has been produced. The conceptual model illustrates the main concepts and relations involved in the maintenance planning problem. It clearly reveals that the central tenet of efficient maintenance planning: preventive maintenance must be performed to a level where the benefits of the maintenance outweigh the costs.

With this in mind, the objective of the statistical analysis and modelling of the data at hand becomes clear: To produce knowledge and to quantify the mechanisms of the conceptual model, in particular the failure process and the effects of preventive maintenance.

By expressing the maintenance planning problem as a conceptual model rather than as a mathematical model, at least initially, a greater flexibility is maintained in the problem solving process. This means that, although the general focus of the

statistical modelling is clear, the specific choice of statistical tools can be conditioned on the data. Similarly, any optimization model which may be applied as a next step can be chosen based on the data and the results of the statistical analysis.

Typically, maintenance optimization methods and models are presented in the literature without these considerations. The properties of the system, the failure process, the mathematical model, and the relevant parameters are given. This means that the models are only applicable in very specific cases. This is exacerbated by the fact that, in order to obtain an optimal solution, the problems are often simplified to the point where the results are of limited use in real life situations.

5.2 Danish Defence data

The Danish Defence has, over the years, developed a high standard in asset tracking and management. This has led to the collection of large databases tracking the history of equipment such as vehicles, ships and aircraft. Even though the data has not been collected for this specific purpose, it is an obvious basis for an analysis to improve maintenance planning. In the course of this project, two of these data sets from the Danish Defence have been analyzed. One covers the maintenance history of 230 tanks, while the other covers the maintenance and usage history 70 fighter aircraft. Both data sets span several years and thousands of observations.

The analyses of both data sets have been carried out using a two step approach. First, a graphical method, Nelson-Aalen plots, is used to investigate the development of the failure intensity over time. This is done for the whole group of systems as well as for different values of covariates, e.g. for different units, and in different time scales. The advantage of using a graphical, non parametric method at this stage is that very few assumptions have to be made beforehand, and a wide range of conclusions can be drawn from the graphs, about the failure intensity, the effects of covariates, and the differences between time scales. This information is useful in itself, and it may be used to make and verify choices and assumptions for further modelling.

The second step of the analyses is performed using multiplicative intensity models with proportional intensities regression. The accuracy of the proportionality assumption is checked on the Nelson-Aalen plots. The regression gives a quantification of the covariate effects found in the graphical analysis. It also makes it possible to test the effects for statistical significance.

While the data quality is generally high, the tank data set in particular contains some errors and inconsistencies. Also, some information regarding the history of the systems between failures has to be inferred because it is not recorded specifically. A lot of these problems can be explained by the fact that the data was collected for accounting purposes and not for reliability analysis and maintenance planning.

Nevertheless, the analyses have generated some interesting and useful results. Generally, the main factor influencing the reliability of the systems is the unit to which they are attached. This has far greater impact than the minor design differences whose effects have also been tested. The analyses have been carried out in

different time scales with roughly similar results, though the differences are a bit more marked in operational time. All of this points to the way the systems are operated and maintained as the main factor influencing their reliability. Thus, while the analyses can not pinpoint the specific causes for the differences, they can serve to focus further investigations into such causes.

In both analyses, the failure intensities remained fairly constant over time. In other words, no wear out effect or reliability growth were found. This means that, since the systems are kept "good as new", the general level of maintenance is high. Whether it is too high, and efficiency may thus be improved by cutting down on preventive maintenance, can not be determined based on the analyses performed here, but it can not be ruled out.

A final product of the analyses is that they have revealed which data are needed for further studies. Often, a few additional pieces of information, or more careful recording of information which is already included in the data, is all that is needed to make the analyses more accurate, and to perform further analyses. As relevant information and advice has been given based on the data already available, a compelling argument has been provided for further data collection.

The method used in the analysis of Danish Defence data has proved very well suited for the given problem and data, thus also validating the problem oriented approach used in selecting which tools and models to use. By selecting methods and models based on the properties of the data sets, while focusing on the overall problem, the maximum amount of information has been extracted from the data.

Appendix A

Analysis of Failure Intensities using Nelson-Aalen plots

Abstract

Based on real life data for tanks from the Danish Army, a statistical analysis of the variation of failure intensity over time is described. A well known graphical method consists of plotting the observed average cumulative number of failures as a function of time (the so called Nelson-Aalen plot). In many practical situations, the exposed population will not be constant over time, but decrease due to replacements, repairs etc. This can lead to a steeper slope for the last part of the Nelson-Aalen plot than would be the case with a constant population. This phenomenon is illustrated with the real life data, and is further investigated for different population developments using simulation.
This report has previously been published as [31] and [29].

A.1 Introduction

The theory of reliability and failure times for systems has traditionally confined itself to the analysis of life times, survival data or times between failures. A repairable system, however, can experience any number of failures in its lifetime, and it is the pattern of these failures that is often of real interest. Thus, in the analysis of failure times for repairable systems, the traditional approach focusing on times between failures often proves insufficient. Instead, one needs to focus on the counting process $N(t)$, where t can be either 'calendar time', measured from a given origin,

but more often indicates a measure of individual system age, or 'operational time'. $N(t)$ denotes the cumulated number of failures a system has experienced since $t = 0$. All these concepts are dicussed thoroughly in [4].

The Nelson-Aalen plot is a non-parametric, graphical approach to the analysis of the failure process for a group of repairable systems. It is based on the Nelson-Aalen estimator for the mean cumulated number of failures per system, MCNF$(t) = M(t)$, defined as $M(t) = E[N(t)]$. The analysis requires a set of observed failure times $t = T_1, T_2, \ldots$, with information about the number of systems which are active at these times. The Nelson-Aalen plot was introduced in [25], and the statistical properties are discussed in [1]. For survival data, it is analogous to the Kaplan-Meier estimator, described in [17]. A thorough treatment of the Nelson-Aalen estimator and its statistical properties can be found in [2].

The significance of $M(t)$ is due to the fact that the slope of the curve, $m(t) = \frac{dM(t)}{dt}$, defines the intensity or ROCOF (rate of occurrence of failures) at time t. Thus, a straight line $M(t)$ implies a constant ROCOF, while an increasing or decreasing slope implies an increasing or decreasing ROCOF. In other words, the slope of the Nelson-Aalen plot is a non-parametric ROCOF estimate. Therefore, the plot can be used to identify trends in the development over time of the failure intensity in a data set.

The Nelson-Aalen estimator is defined as follows

$$\widehat{M}(t) = \sum_{s \leq t} \frac{\Delta N(s)}{V(s)} \tag{A.1}$$

where $V(t)$ is the number at risk, i.e. the number of systems active, at age t, and $\Delta N(t)$ is defined as

$$\Delta N(t) = \begin{cases} 1, & \text{if } t \in \{T_i\} \\ 0, & \text{otherwise} \end{cases}$$

From the definition, it is clear that the Nelson-Aalen plot is an increasing step-function, which jumps only at the times of failure. Therefore, it is sufficient to know the set of jump points

$$\{T_i, \widehat{M}(T_i)\}$$

in order to produce the plot. These points can be calculated recursively using the formula

$$\widehat{M}(T_0) = 0$$

$$\widehat{M}(T_i) = \widehat{M}(T_{i-1}) + \frac{1}{V(T_i)}, i \geq 1 \tag{A.2}$$

A.2 Case study

Figure A.1 shows a Nelson-Aalen plot of failure data from a fleet of 230 Danish Army tanks. Shown together with the Nelson-Aalen plot is a plot showing the number at

risk as a function of time. The time scale used is operational time, that is, individual system age.

The plot shows a typical development of the number at risk. Since this type of tank has been introduced into the service over a period of time, the number at risk is a decreasing function of operational time. Only the tanks introduced early have experienced long service life. This means that the right hand side of the plot is based on relatively few observations from an ever decreasing number at risk.

The first part of the Nelson-Aalen plot shows a linear trend, corresponding to a constant ROCOF. For large values of system age t, however, there is a change in the slope of the plot, which seem to indicate a growing failure intensity for older tanks. This could lead to the conclusion that the older tanks were nearing an age where major overhaul or replacement should be considered.

It is however, important to note that this part of the plot is based on observations on relatively few systems, and that their number is decreasing further. Because of the properties of the Nelson-Aalen estimator under these circumstances, as well as certain characteristics of the data set, such a conclusion could be premature. The rest of this article is dedicated to a further investigation of these phenomenon.

A.3 The effects of decreasing number at risk

The following considerations are illustrated in figure A.2. Consider two systems, A and B, from a data set like the one described above. System A is removed from risk at age $t = T_{dA}$, and system B experiences a failure at $t = T_{fB}$. Thus, using the notation from (A.1),

$$V(T_{dA}^{+}) = V(T_{dA}^{-}) - 1$$

and

$$\widehat{M}(T_{fb}) = \widehat{M}(T_{fb}^{-}) + \frac{1}{V(T_{fb})}.$$

Suppose that T_{dA} and T_{fB} are numerically close, so that nothing else happens between the two incidents. Now, suppose that B fails before A is removed, i.e. $T_{fB} < T_{dA}$. Then

$$\widehat{M}(T_{fB}) = \widehat{M}(T_{fB}^{-}) + \frac{1}{V(T_{dA}^{-})}.$$

In the opposite case, however, with $T_{fB} > T_{dA}$, we get

$$\widehat{M}(T_{fB}) = \widehat{M}(T_{fB}^{-}) + \frac{1}{V(T_{dA}^{-}) - 1}.$$

This means that the exact value of B's failure time determines not only where on the t axis the step in the Nelson-Aalen plot occurs, but also the size of the step, the

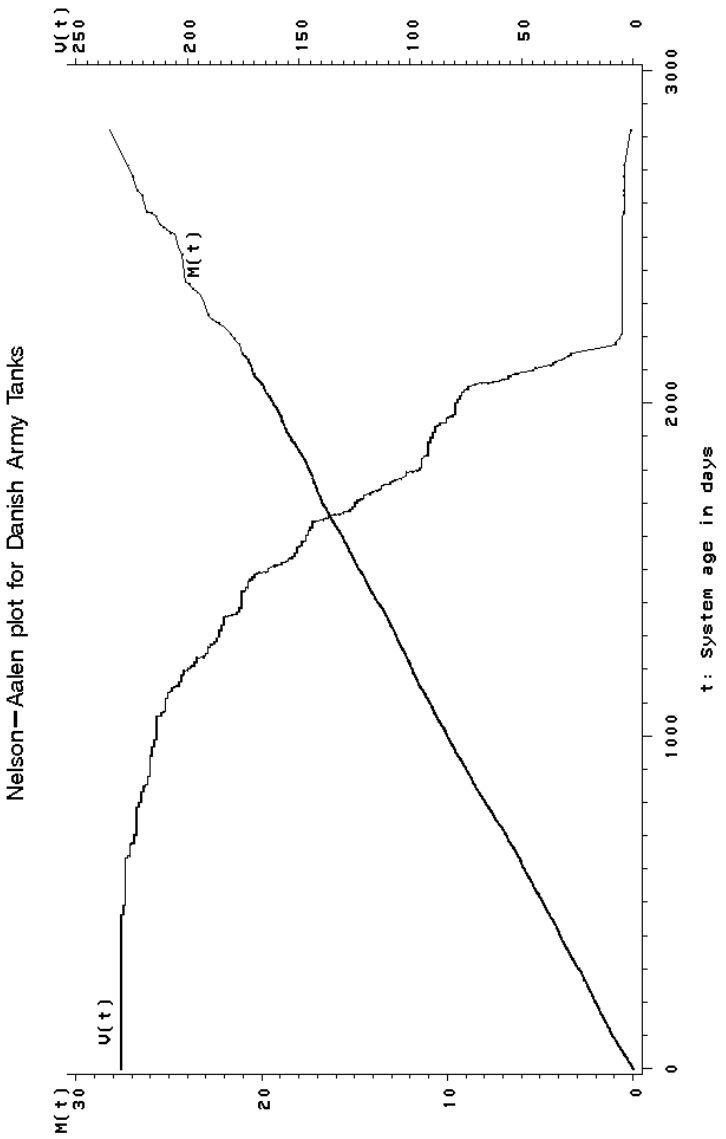

Figure A.1: Nelson-Aalen plot for Danish Army tanks.

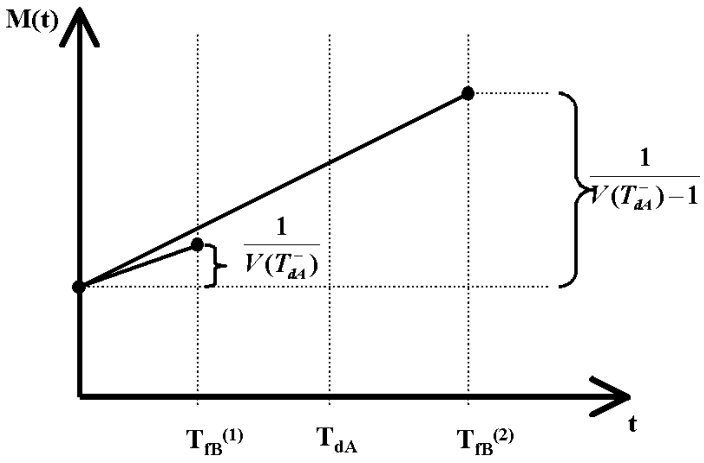

Figure A.2: The effect of a decrease in number at risk.

latter case resulting in a steeper slope on this part of the curve. For large numbers at risk, the difference between the two cases is small, but for a small number at risk, it will be quite significant.

Looking at figure A.1, where events of failure and withdrawal from risk are interspersed all along the age t axis, it is obvious that situations like the one described above will occur many times. This marks the fundamental difference from a situation where the number at risk is constant, that is, where all the withdrawals happen at or near the same age. With the decreasing number at risk, variation in the data will ensure that both cases, $T_{fB} < T_{dA}$ and $T_{fB} > T_{dA}$, occur with about the same probability. A constant number at risk, on the other hand, means that the failures always come first, i.e. $T_{fB} < T_{dA}$.

If we were to compare two sets of data, one with decreasing and one with a constant number at risk, but with identical ROCOF, it is now clear why we would expect to get larger values of $\widehat{M}(t)$ with the decreasing number at risk, especially for larger t: The cumulative nature of the Nelson-Aalen estimator compounds the effect of all previous large steps, while the ever smaller number at risk makes for ever larger steps.

The mechanisms described here are investigated further in the simulation study presented in the following.

A.4 Simulation study

To provide data sets similar in nature to the tank data set, a Monte Carlo simulation of a fleet of 230 repairable systems has been conducted. Each repairable system is modelled as a system of 100 similar components, connected in series, so that the system fails whenever a component fails. Every time a component fails, the time is recorded in the data set, and the faulty component is replaced with a new one. The life time distributions for the components are chosen so that the ROCOF for the systems is kept within the same order of magnitude as the tanks, but apart from that, the simulation is not intended to model the behaviour of the tanks in any detail.

Figure A.3 shows Nelson-Aalen plots for 9 simulated data sets, plotted in the same graph for comparison. All component life times follow an exponential distribution with mean value 10^4. One set (thick line) has all 230 systems running from start ($t = 0$) to finish ($t = 2500$), while the numbers at risk for the remaining 8 are decreasing with increasing age t as shown on the plot.

A system like those simulated here has a constant ROCOF vith a value of 10^{-2}. Therefore, we would expect the Nelson-Aalen plots to follow a straight line with a slope of 10^{-2}. The plot of the data from the experiment with a constant number at risk coincides very well with this line. The 8 plots with decreasing numbers at risk show greater deviance from the line. This would be expected, since they are based on fewer observations.

More noteworthy is the fact that only one of these plots lies below, though fairly close to, the model line. The other 7 plots with decreasing numbers at risk all end up above the line. Furthermore, for larger t, more plots run above the model line, and their distance from the line tends to increase. All of this corresponds neatly with the discussion above.

Numerous other simulations have been carried out, involving different ROCOF functions and number at risk developments over t. All results seem to confirm the theory that with identical ROCOF, decreasing numbers at risk lead to steeper Nelson-Aalen plots than constant numbers at risk. No clear conclusions emerge, however, about the effect of specific ROCOFs or developments in the number at risk.

A.5 The effect of an erroneous estimate of the number at risk

There are additional reasons why the apparent rise in failure intensity for the data illustrated in figure A.1 should be considered with some suspicion. The data set is marred by the fact that the exact time at which the individual tanks began active service is not known. This means that the origin of the time scale must be set to the time of first failure, thus disregarding the time prior to the first failure. This in turn means that we tend to underestimate the number at risk at certain times, specifically where the number is is decreasing. Here, because we underestimate the age of the systems, the drop in the number at risk becomes premature.

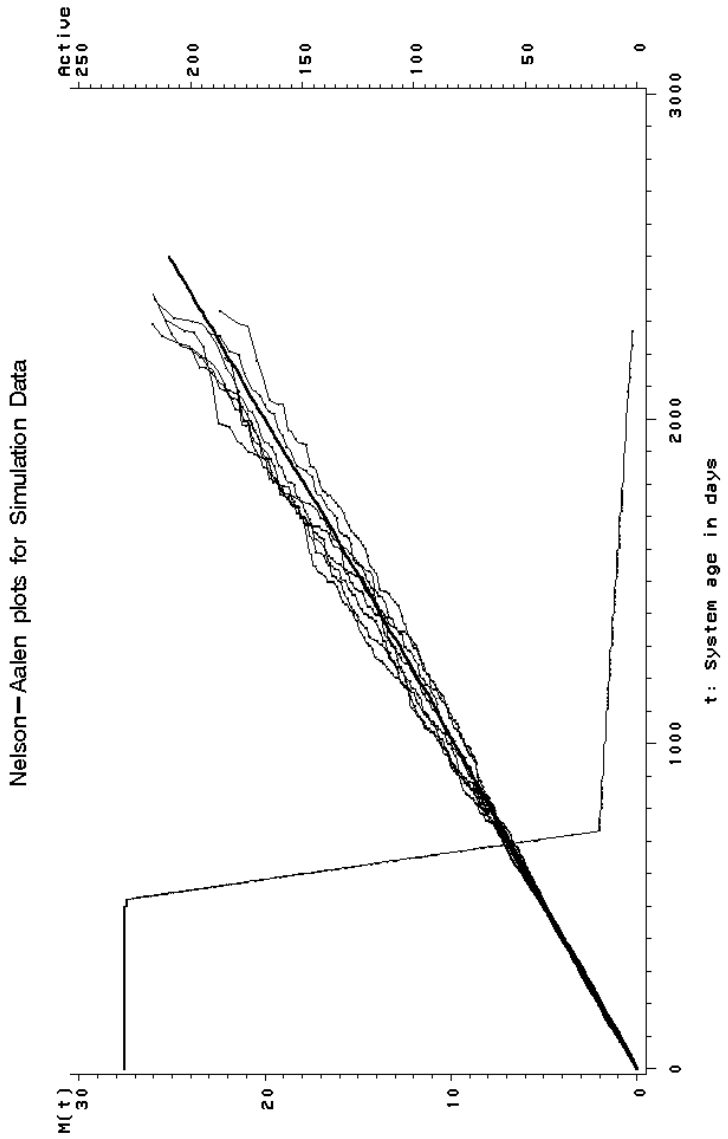

Figure A.3: Nelson-Aalen plots for 9 simulated data sets.

A glance at (A.2) shows why this is significant. Underestimating the number at risk leads to an overestimation of the step size, and thus to the slope of the Nelson-Aalen plot becoming too steep.

The error caused by this tendency to underestimate the number at risk is aggravated by the effects of decreasing numbers at risk. In the notation used earlier, underestimating the age at which a system is withdrawn from risk means that we tend to underestimate T_{dA}. Compared to a situation where T_{dA} is correct, this increases the probability that a failure happens after T_{dA}, i.e. $T_{fB} > T_{dA}$, the case that leads to the larger step. With the probability of larger steps thus increased, the Nelson-Aalen plot will tend to show even steeper slopes. Hence, even small underestimations of the number at risk can lead to significant overestimations of the ROCOF.

A.6 Conclusion

In the analysis of failure intensities for repairable systems, we are often interested in investigating the development of the ROCOF as a function of system age t. This often gives us data sets where the numbers at risk decrease as t grows. The Nelson-Aalen plot, when applied to such data, has a tendency to show increasing ROCOF for large t. An explanation is put forth in this article, where the tendency is explained by comparing the situation where failure times and withdrawals from risk are interspersed along the t axis, i.e. decreasing numbers at risk, with a situation with a constant number at risk, where withdrawals are confined to the end of the t axis. Monte Carlo simulation experiments have been carried out confirming that the Nelson-Aalen plot does indeed have an "upward bias" when applied to data with decreasing numbers at risk. Furthermore, it is shown how even small errors in the estimation of the number at risk can aggravate this effect, leading to significant errors in estimating the ROCOF.

Appendix B

Modelling the failure process for a population of repairable systems using multiplicative intensity models

Abstract

This paper describes a case study on modelling the failure intensity for a population of repairable systems using multiplicative intensity models. The multiplicative intensity model is known to be a very versatile model for counting processes, integrating the modelling of intensity, censoring, and the influence of explanatory variables. Maximum partial likelihood estimation is used to estimate the effect of various explanatory variables. Tests for goodness of fit for the model are presented, and methods for selecting which explanatory variables to include in the model are explored. Furthermore, some practical problems with obtaining and modelling repair times, risk exposure times etc. based on maintenance reports are discussed.
This report has previously been published as [30].

B.1 The Case

This paper is concerned with the analysis of a set of maintenance data for a group of 230 Danish Army vehicles. The purpose of the analysis is to identify factors

which impact reliability and to estimate their effect. Earlier studies using Nelson-Aalen plots ([29], [26]) have been conducted with the data to investigate how failure intensity varies with time, so in this study we wish to concentrate on other factors.

Each time a vehicle is brought in for inspection and/or repairs, a work report is generated at the garage. The data set contains data taken from these work reports. It was collected over a period of several years and contains around 8500 observations, about 6400 of which involve repairs. Each observation in the data set contains the following information:

- Identification of the vehicle.
- Time of the beginning of the work.
- Time of the end of the work.
- Unit to which the vehicle was attached at the time.
- Production year of the vehicle.
- Type of work performed (repair, inspection, or both).

There are several issues with the data which should be taken into account. First of all, it should be noted that the vehicles are repairable systems, as they can experience more than one failure in their lifetime. This means that survival analysis or life-time models can not be directly applied.

Perhaps most noticeable is the fact that data are only recorded at times of failure (repair) or inspection. This means that important information, most of it pertaining to the estimation of the number of vehicles at risk at a given time, is not directly available. For example, it is not known exactly when a vehicle enters service or is transferred to another unit, as nothing indicates that this has happened until the unit experiences a failure. Thus, as the number at risk, or *risk set*, cannot be updated until the following failure, it will be inaccurately estimated. In fact, though we define that a repair means that a failure has occured immediately before, the exact failure time itself is only known in the sense that it can be assumed to be equal to the time when repair is begun.

B.1.1 Preliminary treatment

The analysis will be performed using the SAS system. To be able to work with the data set in SAS, it is transformed to the so-called counting process style. Here, each observation corresponds to a period of time between incidents, these being either entry into the risk set, a repair or inspection, or withdrawal from the risk set. The data set contains the following variables:

- T1: The beginning of the time interval.
- T2: The end of the time interval, and thus the time where a failure may occur.

- EVENT: Takes value 1 if a failure happened at time T2, otherwise 0.
- Y1-Y5: 0,1 indicator variables for each of 5 possible production years of the vehicle. These variables are constant for each vehicle.
- U1-U8: 0,1 indicator variables for each of 8 units. These may vary with time.
- TSI: Time since last inspection. To keep this variable at around the same order of magnitude as the others, it is measured in months. Naturally, this variable also varies with time.

In fact, two data sets have been created, each using a separate time scale for T1 and T2. Originally, times are registered in calendar time, that is, by date and time. Time may, however, also be measured in operational time, where times are measured from a vehicles first entry into the data set. While this may be considered a more relevant measure, especially if the reliability of a vehicle is closely related with its age, it also presents a problem: As already mentioned, we do not know the exact time when a vehicle enters service, and will therefore have to use the time when it first shows up in the data set.

B.2 The multiplicative intensity model

For each system in the population we define a *counting process* $N_k(t)$ which counts the number of failures experienced by the k^{th} system in the interval $[0;t]$. The counting process for the entire population of K systems is defined as

$$N(t) = \sum_{k=1}^{k=K} N_k(t) \tag{B.1}$$

For a counting process $N(t)$ there exists a *compensator* $\Lambda(t)$ so that $N(t) = \Lambda(t) + M(t)$ where $M(t)$ is a *martingale*, meaning, in essence, a pure noise process. The martingale property may be written as

$$E\left[dM(t) \mid \mathcal{F}\left(t^-\right)\right] = 0 \tag{B.2}$$

where $\mathcal{F}(t^-)$ is the *filtration* or *history*, that is, all observations up to time t. The *intensity* process $\lambda(t)$ for the counting process $N(t)$ is defined by

$$E\left[dN(t) \mid \mathcal{F}\left(t^-\right)\right] = \lambda(t)\,dt \tag{B.3}$$

Note that in (B.3), the history of the process $\mathcal{F}(t^-)$ only appears on the right-hand side of the equation, meaning that the intensity is independent of the past history of the process and only depends on t.

Because

$$\Lambda(t) = \int_0^t \lambda(u)\,du \tag{B.4}$$

the compensator $\Lambda(t)$ is sometimes called the integrated or cumulated intensity.

Under the multiplicative hazards, or, more accurately, multiplicative intensity model for the failure intensity of the k^{th} individual in a population of systems can be written as

$$\lambda_k(t) = Y_k(t)\,\alpha_k(t) \tag{B.5}$$

where $Y_k(t)$ assumes the values 0 or 1 indicating whether the system is at risk at time t and $\alpha_k(t)$ is the failure intensity for the system, given that it is under risk at time t. For $N(t)$, assuming that $\alpha_k(t) = \alpha(t)$ for all k, the intensity then becomes

$$\lambda(t) = Y(t)\,\alpha(t) \tag{B.6}$$

where $Y(t) = \sum_{k=1}^{k=K} Y_k(t)$ now gives the size of the *risk set* just before time t. The theory of counting processes, martingales and intensities as well as the multiplicative intensity model are extensively covered in [2].

B.2.1 Proportional intensities regression

Expanding the multiplicative intensities model to allow the intensity to depend on other covariates besides time t, we may expand $\alpha_k(t)$ into

$$\alpha_k(t) = \alpha_0(t)\exp(\beta_1 z_{1k}(t) + \beta_2 z_{2k}(t) + ...) \tag{B.7}$$

where $z_{nk}(t)$ is the value of the n^{th} covariate for the k^{th} individual at time t and β_n is the parameter corresponding to the n^{th} covariate. $\alpha_0(t)$ is a baseline intensity for all covariates $z_{nk} = 0$. This function may be assumed to have a given parametric form, but is more often regarded as unknown, because one is mainly interested in the effect of the covariates.

Because the covariates may vary with time and may depend on the past history of the process, proportional intesities regression provides a way to let the intensity depend on the past history of the process.

B.3 Parameter estimation

The intensity $\lambda_k(t)$ may be interpreted as the probability of a failure at instant t. Conversely, the probability of no failure at t is $1 - \lambda_k(t)$. The full likelihood for the entire population may then, with the parameter vector $\boldsymbol{\theta}$ containing possible parameters for $\alpha_0(t)$, be written as the product integral

$$\mathcal{L}(\beta,\boldsymbol{\theta}) = \prod_k \mathcal{L}_k(\beta,\boldsymbol{\theta}) = \prod_k \prod_t \left((\lambda_k(t,\beta,\boldsymbol{\theta}))^{dN(t)} (1 - \lambda_k(t,\beta,\boldsymbol{\theta}))^{1-dN(t)} \right) \tag{B.8}$$

where $dN(t)$ is 1 if a failure occurs at time t and 0 otherwise. While the likelihood in this form is compact and intuitively appealing, it is not very practical when it comes to deriving actual estimators for the parameters. Assuming for simplicity that the covariates for each system are constant over time, we may write the likelihood as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\theta}) &= \exp\left(-\sum_k \int_0^\infty \lambda_k(u,\boldsymbol{\beta},\boldsymbol{\theta})\,du\right)\prod_k\prod_{i=1}^{n_k}\lambda_k(t_{ik},\boldsymbol{\beta},\boldsymbol{\theta}) \\ &= \exp\left(-\sum_k \int_0^\infty Y_k(u)\,\alpha_0(u,\boldsymbol{\theta})\,du\exp(\boldsymbol{\beta}\mathbf{z}_k)\right) \\ &\qquad \exp\left(\sum_k n_k\boldsymbol{\beta}\mathbf{z}_k\right)\prod_k\prod_{i=1}^{n_k}\alpha_0(t_{ik},\boldsymbol{\theta})\end{aligned} \tag{B.9}$$

This form still presents some problems, and most of these concern $\alpha_0(t)$. If $\alpha_0(t)$ is assumed to have a parametric form, (B.9) may become quite complex and hence usuitable for parameter estimation in practice. If on the other hand, $\alpha_0(t)$ is unknown, one has to figure out a way to factor it out of the likelihood.

B.3.1 Partial likelihood

The theory of partial likelihood provides a way to factor out nuisance parameters of secondary or no importance. It was first introduced by Cox in [11] and expanded to cover multiple failures by Prentice, Williams and Peterson in [33]. The partial likelihood is not a likelihood in the usual sense that it is proportional to the conditional or marginal probability of any observation. Cox [12] shows, however, that maximum partial likelihood estimates are consistent and asymptotically normally distributed.

The idea is to construct the likelihood so that it contains all or most of the relevant information without containing the unimportant parameters. This is done by dividing the observed information into a sequence

$$(X_1,S_1,X_2,S_2,\ldots,X_m,S_m) \tag{B.10}$$

The full likelihood for this sequence is

$$\prod_{j=1}^{m} f_{X_j|X^{(j-1)},S^{(j-1)}}\left(x_j|x^{(j-1)},s^{(j-1)};\theta\right)\prod_{j=1}^{m} f_{S_j|X^{(j)},S^{(j-1)}}\left(s_j|x^{(j)},s^{(j-1)};\theta\right) \tag{B.11}$$

where $x^{(j)}=(x_1,\ldots,x_j)$ and $s^{(j)}=(s_1,\ldots,s_j)$. The second product is the partial likelihood based on S in the sequence (B.10). The factorization of the full likelihood could be done in several different ways. To make the partial likelihood useful in practice, however, some guidelines have to be observed. The omitted factors (the X's) should not contain important information about the parameters of interest and, likewise, the partial likelihood should not contain unimportant parameters.

The data available in our case can be seen as a sequence similar to (B.10). Let S_{ik} be the event that the i^{th} failure for system k occurs at t_{ik}, and let X_{ik} be the event

that specifies covariate, counting process and censoring information in the interval $[t_{(i-1)k}, t_{ik}]$. The partial likelihood based on S will then be

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}) &= \prod_k \prod_{i=1}^{n_k} P\left\{s_{ik} | x^{(i)}, s^{(i-1)}; \boldsymbol{\beta}\right\} \\ &= \prod_k \prod_{i=1}^{n_k} \frac{\alpha_0\left(t_{ik}\right) \exp\left(\boldsymbol{\beta}\mathbf{z}_k\left(t_{ik}\right)\right)}{\sum_l Y_l\left(t_{ik}\right) \alpha_0\left(t_{ik}\right) \exp\left(\boldsymbol{\beta}\mathbf{z}_l\left(t_{ik}\right)\right)} \\ &= \prod_k \prod_{i=1}^{n_k} \frac{\exp\left(\boldsymbol{\beta}\mathbf{z}_k\left(t_{ik}\right)\right)}{\sum_l Y_l\left(t_{ik}\right) \exp\left(\boldsymbol{\beta}\mathbf{z}_l\left(t_{ik}\right)\right)} \end{aligned} \tag{B.12}$$

where $\mathbf{z}_l\left(t_{ik}\right)$ is the covariate vector for the l^{th} system at the time of the i^{th} failure of the k^{th} system.

We now have a partial likelihood which includes only the covariate parameters and the information that we have available in the data set, while excluding the baseline, time dependent intensity, which we are less interested in, and information for times between failures, which is unavailable to us.

B.4 Tests in the model

Tests exist both for the hypothesis $\boldsymbol{\beta} = \mathbf{0}$, meaning that none of the covariates have any effect, and for linear hypotheses of the form $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$ where $\mathbf{A}$ is a coefficient matrix and $\mathbf{c}$ is a vector of constants. The Wald statistic

$$T_W = \left(\mathbf{A}\widehat{\boldsymbol{\beta}} - \mathbf{c}\right)^T \left(\mathbf{A}\widehat{\mathbf{V}}\left(\widehat{\boldsymbol{\beta}}\right)\mathbf{A}^T\right)^{-1} \left(\mathbf{A}\widehat{\boldsymbol{\beta}} - \mathbf{c}\right) \tag{B.13}$$

where

$$\widehat{\mathbf{V}}\left(\widehat{\boldsymbol{\beta}}\right) = -\left(\frac{\partial^2 \log \mathcal{L}(\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^2}\right)^{-1} \tag{B.14}$$

is the estimated covariance matrix of $\widehat{\boldsymbol{\beta}}$, is, under the assumption $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$, assymptotically $\chi^2(n)$ distributed. The n degrees of freedom corresponds to the rank of $\mathbf{A}$. Since the null hypothesis $\boldsymbol{\beta} = \mathbf{0}$ is a special case of the linear hypothesis, the Wald test applies here as well (B.15). Other test statistics for the null hypothesis are the likelihood ratio test (B.16) and the score test (B.17), both of which also assymptotically follow a $\chi^2(n)$ distribution.

$$T_W^0 = \widehat{\boldsymbol{\beta}}^T \left(\widehat{\mathbf{V}}\left(\widehat{\boldsymbol{\beta}}\right)^{-1}\right) \widehat{\boldsymbol{\beta}} \tag{B.15}$$

$$T_{LR}^0 = 2\left(\log \mathcal{L}(\widehat{\boldsymbol{\beta}}) - \log \mathcal{L}(\mathbf{0})\right) \tag{B.16}$$

$$T_S^0 = \left(\frac{\partial \log \mathcal{L}(\mathbf{0})}{\partial \boldsymbol{\beta}}\right)^T \left(-\frac{\partial^2 \log \mathcal{L}(\mathbf{0})}{\partial \boldsymbol{\beta}^2}\right)^{-1} \left(\frac{\partial \log \mathcal{L}(\mathbf{0})}{\partial \boldsymbol{\beta}}\right) \tag{B.17}$$

B.5 Results

The models and tests described above are implemented in the SAS procedure PHREG. The procedure was used to analyze the data set with times indicated as calendar time as wells as operational time. Since no major differences were found between the results for the two time scales, however, we will concentrate on the results for the data in calendar time.

Initially, the procedure has been run with only production year or unit as covariates. In each case, one variable is left out. The reason the variable for one production year and one unit have to be left out is that since the year and the unit variables sum to 1, each variable is completely defined by the other variables of the same type. Thus vehicles with Y1=1 and U1=1 are selected as having the baseline failure intensity.

The tables below show the maximum partial likelihood estimates for the parameters as well as the Wald statistic for the hypothesis $\widehat{\beta_n} = 0$ for each parameter and the value of $p = P\left[T_W < \chi^2(1)\right]$. Finally, the risk relative to the baseline intensity, a more intuitive measure of the effect of the covariate, is given.

Variable n	$\widehat{\beta_n}$	T_W	$p = P\left[T_W < \chi^2(1)\right]$	Risk ratio
Y2	-0.026143	0.18107	0.6705	0.974
Y3	-0.189343	23.66649	0.0001	0.828
Y4	-0.071887	4.30115	0.0381	0.931
Y5	-0.321824	51.54113	0.0001	0.725

Table B.1: Results with the production year variables only.

The results in table B.1 show that failure intensity varies significantly with production year, though the difference between years 1 and 2 is clearly insignificant. With the lowest failure intensity at 72.5% of the highest, the real difference, if not dramatic, is nevertheless notable.

Variable n	$\widehat{\beta_n}$	T_W	$p = P\left[T_W < \chi^2(1)\right]$	Risk ratio
U2	-0.209786	42.07947	0.0001	0.811
U3	-0.109536	5.97422	0.0145	0.896
U4	-1.187680	253.45249	0.0001	0.305
U5	-0.755206	310.05191	0.0001	0.470
U6	-0.563297	121.47653	0.0001	0.569
U7	-1.065917	321.82624	0.0001	0.344
U8	-1.636261	97.12088	0.0001	0.195

Table B.2: Results with the unit variables only.

In contrast to the previous table, the differences in failure intensity between different units, shown in table B.2, are both statistically significant and quite dramatic in real terms, with the variations up to a factor of 5 failures in unit 1 for every one in unit 8.

Table B.3 shows the results when the procedure is run on the full model, that is, including all the covariates except Y1 and U1.

Variable n	$\widehat{\beta_n}$	T_W	$p = P\left[T_W < \chi^2(1)\right]$	Risk ratio
Y2	0.061632	0.98667	0.3206	1.064
Y3	0.041025	0.95744	0.3278	1.042
Y4	0.170207	18.44116	0.0001	1.186
Y5	-0.085855	3.35338	0.0671	0.918
U2	-0.201452	38.35142	0.0001	0.818
U3	-0.169986	11.77252	0.0006	0.844
U4	-1.235760	267.00090	0.0001	0.291
U5	-0.759818	268.33169	0.0001	0.468
U6	-0.584860	128.58018	0.0001	0.557
U7	-1.067821	320.70163	0.0001	0.344
U8	-1.764976	110.93598	0.0001	0.171
TSI	-0.013931	51.34919	0.0001	0.986

Table B.3: Results with the full model.

When comparing these results with the first two tables, one notices that while the unit effect is roughly the same as before, the estimated effect of production year is quite different. Specifically, only year 4 now seems to have a significant effect, although it was not the most significant effect in the earlier model. Also, the estimated relative effects of the years have changed considerably.

The reason for these discrepancies is that the different years are not distributed evenly among the units. For example, units 1 and 2 have a large proportion of year 1 vehicles, while unit 3 has many with production year 4. What happens when the model is fitted with both types of covariates is that it must "assign the blame" for high failure intensities to either unit or production year. In this case, it seems that unit 1 and 2 are blamed for their relatively high intensity, while the high intensity in unit 3 is blamed on the production year of its vehicles.

This phenomenon, known as multicollinearity, makes it hard to determine the effect of any one covariate. The behavior we have seen here, with the estimated effect of one explanatory variable changing dramatically when another is included, is typically associated with multicollinearity. In a case such as this, where we can not influence how data is generated, e.g. change how the vehicles are distributed among the units, there is very little we can do statistically to reliably determine the effect of the individual covariates. The necessary information is simply not there.

It is, however, striking that the unit effect "survives" the inclusion of the other variables, in the sense that the estimated relative risks are numerically close and still significant. In the absence of any additional information, for example about important physical differences between vehicles from different years, we must conclude that the results point to the unit as an important influence on the reliability of the vehicle, while the observed effect of the production year is less clear and may well

result from the multicollinearity between the two covariates.

The time since inspection, which is also included here, has a small but significant effect. This result has been confirmed by fitting the model with this covariate alone, as well as with both of the other types of covariates. Truly noteworthy is, of course, the fact that the parameter is negative, meaning that the failure intensity actually decreases with time after an inspection. It is tempting to conclude that inspections are damaging reliability and should be limited or avoided altogether. There are, however, other explanations for this result. Failures found at an inspection may be repaired later under a separate work order, thus counting as a failure shortly after inspection, or inspection may have been carried out as a precaution shortly before the vehicle is to be heavily used, thus resulting in more failures.

B.6 Conclusion

We have modelled maintenance data for a group of Army vehicles with a model with multiplicative intensity and proportional intensity regression. Using the theory of partial likelihood, we have been able to estimate the effect of explanatory covariates on the failure intensity of the vehicles and to test which of these effects are significant. The failure intensity of a vehicle was found to be heavily influenced by the unit to which it is attached, while the time since last inspection and the production year of the vehicle were found to be of lesser importance.

Apart from the problems with the data concerning the proper delineation of the risk set, which were already known from the outset, we have encountered problems with multicollinearity between some covariates, which makes it difficult to separate the effects of the individual covariates. These problems have more to do with the data than the method, though, and more informative solutions would require additional information. Thus, the model and estimation method presented here seem to provide important and useful tools for modelling and analyzing maintenance data and repairable systems in general.

Bibliography

[1] O. O. Aalen. *Statistical inference for a family of counting processes.* PhD thesis, Institute of Mathematical Statistics, University of Copenhagen, 2100 Copenhagen Ø, DK, 1976.

[2] Per Kragh Andersen, Ørnulf Borgan, Richard D. Gill, and Niels Keiding. *Statistical Models Based on Counting Processes.* Springer Series in Statistics. Springer Verlag New York, Inc., New York, 1993.

[3] S. Apeland and T. Aven. Risk based maintenance optimization: foundational issues. *Reliability Engineering and Systems Safety*, 67:285–292, 2000.

[4] Harold E. Ascher and Harry Feingold. *Repairable Systems Reliability: Modelling, Inference, Misconceptions and their Causes.* Marcel Dekker, New York, 1984.

[5] Harold E. Ascher and Christian K. Hansen. Spurious exponentiality observed when incorrectly fitting a distribution to nonstationary data. *IEEE Transactions on Reliability*, 47(4):451–459, December 1998.

[6] Harold E. Ascher and Khairy A. H. Kobbacy. Modelling preventive maintenance for deteriorating repairable systems. *IMA Journal of Mathematics Applied in Business & Industry*, 6:85–99, 1995.

[7] Frans A. Boshuizen. A replacement model with general age-dependent failure rates. *Journal of Statistical Planning and Inference*, 59:213–228, 1997.

[8] Siang-Ying Choy, John R. English, Thomas L. Landers, and Li Yan. Collective approach for modeling complex system failures. In *1996 Proceedings Annual Reliability and Maintainability Symposium*, pages 282–286. IEEE, 1996.

[9] A. H. Christer and W. M. Waller. An operational research approach to planned maintenance: Modelling p.m. for a vehicle fleet. *Journal of tho Operational Research Society*, 35(11):967–984, 1984.

[10] Jasper L. Coetzee. The role of NHPP models in the practical analysis of maintenance failure data. *Reliability Engineering and Systems Safety*, 56:161–168, 1997.

[11] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society Series B*, 34:187–220, 1972.

[12] D. R. Cox. Partial likelihood. *Biometrika*, 62(2):269–276, 1975.

[13] J. Crocker and U. D. Kumar. Age-related maintenance versus reliability centered maintenance: a case study on aero-engines. *Reliability Engineering and Systems Safety*, 67:113–118, 2000.

[14] Safaai Deris, Sigeru Omatu, Hiroshi Ohta, Lt. Cdr Shaharudin Kutar, and Pathiah Abd Samat. Ship maintenance scheduling by genetic algorithm and constraint-based reasoning. *European Journal of Operational Research*, (112):489–502, 1999.

[15] R. Guo and Charles E. Love. Statistical analysis of an age model for imperfectly repaired systems. *Quality and Reliability Engineering International*, 8:133–146, 1992.

[16] R. Guo and Charles E. Love. Simulating nonhomogeneous poisson processes with proportional intensities. *Naval Research Logistics*, 41(4):507–522, 1994.

[17] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete estimations. *Journal of the Amrican statistical Association*, 53:457–481, 1958.

[18] Dimitri Kececioglu. *Reliability Engineering Handbook*, volume 1. Prentice Hall PTR, Upper Saddle River, NJ 07458, 1991.

[19] Khairy A. H. Kobbacy, Bahir B. Fawzi, David F. Percy, and Harold E. Ascher. A full history proportional hazards model for preventive maintenance scheduling. *Quality and Reliability Engineering International*, 13:187–198, 1997.

[20] Khairy A. H. Kobbacy, Nathan C. Proudlove, and Murray A. Harper. Towards an intelligent maintenance optimization system. *Journal of the Operational Research Society*, 46:831–853, 1995.

[21] Jan Terje Kvaløy and Bo Henry Lindqvist. TTT-based tests for trend in repairable systems data. *Reliability Engineering and Systems Safety*, 60:13–28, 1998.

[22] C. E. Love, Z. G. Zhang, M. A. Zitron, and R. Guo. A discrete semi-Markov decision model to determine the optimal repair/replacement policy under general repairs. *European Journal of Operational Research*, 125:398–409, 2000.

[23] Jens Lund. Sampling bias in population studies - how to use the lexis diagram. *Scandinavian Journal of Statistics*, 27(4):589–604, 2000.

[24] Christian Max Møller. Introduktion til punktprocesser. Arbejdsrapport A-7/1998, Forsvarets Forskningstjeneste, 1998.

[25] W. Nelson. Graphical analysis of system repair data. *Journal of Quality Technology*, 20:24–35, 1988.

[26] Thomas Espelund Pedersen. Grafisk analyse af fejlintensiteter for kampvogne. Forskningsrapport F-27/2000, Forsvarets Forskningstjeneste, 2000.

[27] Thomas Espelund Pedersen. Analyse af fejlintensiteter for F-16. Forskningsrapport F-34/2002, Forsvarets Forskningstjeneste, 2002.

[28] Thomas Espelund Pedersen. Analyse af fejlintensiteter for kampvogne. Forskningsrapport F-06/2002, Forsvarets Forskningstjeneste, 2002.

[29] Thomas Espelund Pedersen. Analysis of failure intensities using Nelson-Aalen plots. Arbejdsrapport A-2/2002, Forsvarets Forskningstjeneste, 2002.

[30] Thomas Espelund Pedersen. Modeling the failure process for a population of repairable systems using multiplicative intensity models. Arbejdsrapport A-3/2002, Forsvarets Forskningstjeneste, 2002.

[31] Thomas Espelund Pedersen and Poul Thyregod. Analysis of failure intensities using Nelson-Aalen plots. In *Symposium i Anvendt Statistik 2000*, 2000.

[32] David F. Percy, Khairy A. H. Kobbacy, and Bahir B. Fawzi. Setting preventive maintenance schedules when data are sparse. *International Journal of Production Economics*, 51:223–234, 1997.

[33] R. L. Prentice, B. J. Williams, and A. V. Peterson. On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379, 1981.

[34] Kjell Sandve and Terje Aven. Cost optimal replacement of monotone, repairable systems. *European Journal of Operational Research*, (116):235–248, 1999.

[35] Philip A. Scarf. On the application of mathematical models in maintenance. *European Journal of Operational Research*, (99):493–506, 1997.

[36] Christophe Vandeschrick. The lexis diagram, a misnomer. *Demographic Research*, 4:97–124, 2001.